
DIPLOMARBEIT

Frau
Kathleen Wiesner

**Übergangswahrscheinlichkeiten
von SGB III-Arbeitslosigkeit in
SGB II-Arbeitslosigkeit in Sachsen
und deren Bestimmungsfaktoren**

2010

DIPLOMARBEIT

Übergangswahrscheinlichkeiten von SGB III-Arbeitslosigkeit in SGB II-Arbeitslosigkeit in Sachsen und deren Bestimmungsfaktoren

Autorin:

Kathleen Wiesner

Studiengang:

Angewandte Mathematik

Seminargruppe:

Ma05w1

Erstprüfer:

Herr Professor Lindner

Zweitprüfer:

Frau Dr. Antje Weyh

Mittweida, 2010

Kurzfassung

Seit der Einführung der Änderungen der Arbeitslosengeld II-Gesetze im Jahre 2005, gab es wesentliche Veränderungen in der Arbeitsmarktpolitik. Je nach Art und Länge des vorherigen Arbeitsverhältnisses gibt es 2 Arten von Leistungen, welche ein Hilfeebedürftiger erhalten kann. Es wird unterschieden nach Leistungen aus dem Rechtskreis des SGB III¹ und Leistungen aus dem Kreis des SGB II². Es kann darüber hinaus auftreten, dass sowohl SGB III und SGB II Leistungen empfangen werden.

In der nachfolgenden Diplomarbeit soll die Frage erörtert werden, welche Bestimmungsfaktoren maßgeblich dafür verantwortlich sind, dass Menschen aus der SGB III in eine SGB II Arbeitslosigkeit wechseln. Es wird zunächst auf die Datengrundlage eingegangen und die wichtigsten Begriffe erläutert. Nach einer deskriptiven Auswertung der gewonnenen Daten, sowie der Bestimmung der wichtigsten statistischen Kennzahlen, erfolgt die Auswertung mittels Regression.

Die logistische Regression wird vor allem in der Wirtschaft immer beliebter. Ziel ist es zu bestimmen, mit welcher Wahrscheinlichkeit ein Übergang aus dem SGB III- in den SGB II-Bereich stattfindet. Es wird erörtert, welche Variablen maßgeblich einen Einfluss auf die Übergangswahrscheinlichkeit haben. Als Modelle werden die binäre logistische Regression und die ordinale logistische Regression verwendet. Abschließend wird die Güte der Modelle bewertet. Durch die Analyse sollen vor allem die Risikofaktoren eines solchen Modells bestimmt werden.

Zusätzlich erfolgt eine Auswertung mittels einer Panelanalyse. Dieses Modell ist besonders dann sinnvoll, wenn sich im Verlauf der Zeit (Jahre, Monate, Tage) ein Merkmal einer Person ändert. Es werden die Veränderungen in den Jahren 2007 und 2008 untersucht. Betrachtet werden die Modelle, des Pooled Logit Modells, des Fixed Effects Modells und das Random Effects Modells.

Am Ende der Diplomarbeit wird diskutiert, inwieweit die Lösungen praktikabel sind und wie die Entwicklung in dieser Richtung weiter gehen kann. Passend zu den erzeugten Ergebnissen, wurde ein Tool für Stata entwickelt, welches diese reproduziert. Infolgedessen ist es möglich, die Auswertung für andere Bundesländer vorzunehmen.

¹ 3. Sozialgesetzbuch

² 2. Sozialgesetzbuch

I. Inhaltsverzeichnis

Inhaltsverzeichnis	8
Abbildungsverzeichnis	9
Tabellenverzeichnis	12
1 Definitionen	1
1.1 Arbeitslosigkeit	1
1.2 SGB III Arbeitslosigkeit	2
1.3 SGB II Arbeitslosigkeit	3
1.4 Übergänge und Übergangswahrscheinlichkeiten	4
2 Datenanalyse und Verarbeitung	7
2.1 Aufbau des Datensatzes	7
2.2 Episodensplitting	7
2.3 Datenquellen	9
2.4 Verarbeitung des Datensatzes	11
3 Deskriptive Analyse	15
3.1 Allgemeine Betrachtung des Datensatzes	15
3.2 Auswertung nach personengebundenen Merkmalen	16
3.3 Auswertung nach regionalen Merkmalen	20
3.4 Auswertung nach zeitlichen Merkmalen	23
3.5 Auswertung nach arbeitsmarktpolitischen Maßnahmen	25
4 Auswertung mittels logistischer Regression	27
4.1 Grundlagen	27
4.2 Analyse mittels logistischer Regression	33
4.3 Spezialfall ordinale Regression	44
5 Panelanalyse	53
5.1 Fixed Effects Modell	55
5.2 Random Effects Modell	56
5.3 Auswertung	57
6 Zusammenfassung	61
A Ausprägungen der Variablen	63
A.1 Ausprägungen von der Variablen hoechsterab	63
A.2 Ausprägungen von der Variablen hauptagentur	63
A.3 Ausprägungen von der Variablen ieb_wo_gst_num	64
A.4 Ausprägungen von der Variablen ieb_wo_krs_num	64

B Übergänge in den Geschäftsstellen	65
C Programmcode	67
Literaturverzeichnis	89

II. Abbildungsverzeichnis

2.1	Anzahl der Spells vor und nach dem Episodensplitting	8
3.1	Verteilung der SGB III Empfänger bezogen auf das Alter	17
3.2	Verteilung der Übergänge bezogen auf das Alter	17
3.3	Im Mittel übergegangene Personen nach Schulabschlüssen	19
3.4	Im Mittel übergegangene Personen nach Hauptagenturen	21
3.5	Hochschulabschlüsse - Anteil der Übergänge	23
4.1	Alter dargestellt mit der Variablen Übergänge	28
4.2	lineare Regressionsgerade	28
4.3	Verteilung der Residuen	29
4.4	Verlauf der Linkingfunktion	31
4.5	Verlauf der Likelihoodfunktion	31
4.6	Aufteilung von Sachsen in Kreise	35
4.7	Aufteilung von Sachsen in Hauptagenturen	35
4.8	Sensitivität/Spezifität in Abhängigkeit des Cut-Off Wertes	39
4.9	Roc Kurve	40
4.10	Anzahl der Jahre mit Übergängen - Anzahl der Maßnahmen 2007	45

III. Tabellenverzeichnis

1.1	Synonyme	1
1.2	Länge des Leistungsbezuges nach vorheriger Beschäftigungsdauer	3
2.1	Übersicht der verwendeten Variablen	8
2.2	Spells vor und nach dem Episodensplitting	8
2.3	Bitmusterzuordnung der Quellen	10
2.4	Kombinationen der einfließenden Quellen	12
2.5	Spalten für Panelanalyse	13
3.1	Anzahl der Übergänge und Arbeitslosengeld I Bezieher	16
3.2	Übergegangene Personen nach Geschlecht	16
3.3	Übergegangene Personen nach Abschluss	18
3.4	Übergegangene Personen nach Schwerbehindertenstatus	19
3.5	Übergegangene Personen nach Staatsangehörigkeit	20
3.6	Übergegangene Personen nach Hauptagenturen	21
3.7	Übergegangene Personen nach Kreisen	22
3.8	Anzahl der Personen mit mehreren Übergängen	24
3.9	Anzahl der Personen nach dem Zeitraum zwischen den SGB III und SGB II Bezug	24
3.10	Anzahl der Maßnahmen	25
3.11	Anzahl der Nebenbeschäftigungen	26
4.1	Ausprägungen der verwendeten Variablen	33
4.2	Korrelation zwischen den Variablen und den Übergängen	34
4.3	Kodierung der Variablen höchster Abschluss	36
4.4	Referenzkategorien	37
4.5	1. Ergebnis nach Aufnahme der Konstanten	37
4.6	χ^2 -Werte und Freiheitsgrade nach den 1. Berechnungsschritt	38
4.7	Übersicht über die Entwicklung der Güte des Modells	38

4.8	Ergebnisse nach der vollständigen Aufnahme aller Variablen	41
4.9	Interaktionseffekt	43
4.10	Zusammenhang zwischen y und y^*	44
4.11	1. Ergebnis der ordinalen Regression nach Aufnahme der Konstanten . . .	48
4.12	χ^2 -Werte und Freiheitsgrade nach den 1. Berechnungsschritt der ordinalen Regression	48
4.13	Übersicht über die Entwicklung der Güte des Modells beim ologit Modell . .	48
4.14	Gegenüberstellung der beiden Modelle	49
4.15	Ergebnisse nach der vollständigen Aufnahme aller Variablen im ologit Modell	49
4.16	Vergleich der Modelle nach der ologit und logit Regression	50
4.17	Beispielrechnung für die beiden logistischen Modelle	51
5.1	Daten im Wide Format	54
5.2	Daten im Long-Format	54
5.3	Modell für das Fixed Effect Modell	57
5.4	Modell für das Random Effect Modell	58
5.5	Vergleich der beiden Modelle	59
A.1	Ausprägungen von hoechsteraab	63
A.2	Ausprägungen von hauptagentur	63
A.3	Ausprägungen von ieb_wo_gst_num	64
A.4	Ausprägungen von ieb_wo_krs_num	64
B.1	Übergänge und SGB III Personen nach Geschäftsstellen	65

1 Definitionen

Im folgenden Kapitel werden die Grundbegriffe definiert, welche für das weitere Verständnis der Arbeit von Bedeutung sind. Dieses Kapitel beinhaltet die Definitionen der Begriffe SGB II und SGB III Arbeitslosigkeit. Nachfolgende Tabelle verdeutlicht häufige verwendete Synonyme der beiden Leistungsarten.

SGB III Arbeitslosigkeit	SGB II Arbeitslosigkeit
Arbeitslosengeld I	Arbeitslosengeld II
ALG I	ALG II
SGB III Leistungen	SGB II Leistungen
	Grundsicherung für Arbeits-
	suchende
	„Hartz IV“ ³

Tabelle 1.1: Synonyme

1.1 Arbeitslosigkeit

Arbeitslosigkeit ist im Gesetz definiert als [14]:

- Der Arbeitnehmer darf nicht in einem Beschäftigungsverhältnis stehen.
- Er muss sich bemühen, seine Beschäftigungslosigkeit zu beenden.
- Er muss der Vermittlung durch die Agentur für Arbeit zur Verfügung stehen.

Beschäftigungslosigkeit:

Die Person darf nur eine Tätigkeit ausüben, deren Dauer weniger als 15 Stunden pro Woche beträgt. Durch eine geringfügige Beschäftigung⁴ geht der Anspruch auf Leistungen nicht verloren. Auch leichte Überschreitungen der Arbeitszeit führen nicht zum Verlust des Anspruchs auf Arbeitslosengeld I, sofern diese nicht dauerhaft sind. Ebenso sind Ehrenämter erlaubt, insofern diese die Vermittlungsfähigkeit des Arbeitnehmers nicht beeinträchtigt wird.

Eigenbemühungen:

Unter Eigenbemühungen wird verstanden, dass alle Chancen auf eine berufliche Eingliederung genutzt werden. Der Antragsteller hat Verpflichtungen zu erfüllen. Er muss der Agentur für Arbeit zur Verfügung stehen. Auf Verlangen müssen Belege für das

³ umgangssprachlich; Dieser Begriff wird hier keine Verwendung finden, da ihn eine deutlich negative Tendenz anhaftet.

⁴ Das sind Arbeitsverhältnisse mit niedrigem Lohn. Ähnlich ist die kurzfristige Beschäftigung, welche von vornherein auf eine bestimmte Dauer, zum Beispiel 2 Monaten beschränkt ist.

Erbringen der Verpflichtungen vorgelegt werden können, zum Beispiel Nachweise für geschriebene Bewerbungen. Geschieht dies nicht, so kann dem Antragsteller, wegen unzureichender Eigenbemühungen, die Leistung ganz oder teilweise gekürzt werden.

Zumutbarkeit:

Der Arbeitnehmer muss bereit sein, jede für ihn zumutbare⁵ Arbeit anzunehmen. Ebenso hat der Antragsteller an beruflichen Eingliederungen oder Weiterbildungen teilzunehmen.

Verfügbarkeit:

Der Leistungsempfänger muss örtlich sowie zeitlich für die Agentur für Arbeit zur Verfügung stehen.

- Mitteilungen von der Agentur für Arbeit müssen zur Kenntnis genommen werden können.
- Es kann verlangt werden, dass jederzeit Kontakt mit der Agentur für Arbeit oder einen zukünftigen Arbeitnehmer aufgenommen werden kann.
- Es muss zu jeder Zeit möglich sein, eine Arbeit zu beginnen oder eine Eingliederungsmaßnahme zu besuchen.

1.2 SGB III Arbeitslosigkeit

1.2.1 Voraussetzungen

Arbeitslosigkeit, wie oben definiert, ist eine Voraussetzung für den Empfang von SGB III Leistungen. Es muss eine Anwartschaftszeit⁶ erfüllt sein. Bei der Regelanwartschaft beträgt diese 12 Monate. In den letzten 5 Jahren muss der Antragsteller mindestens 12 Monate einer versicherungspflichtigen Beschäftigung nach gegangen sein. Zu der Anwartschaftszeit hinzu gezählt werden: Wehr oder Zivildienst, Kindererziehung⁷, Bezug von Rente aufgrund voller Erwerbsminderung, Zeiträume in denen eine freiwillige Weiterversicherung vorlag oder einer Beschäftigung im Ausland nachgegangen wurde. Dabei gilt, dass zuvor jeweils eine versicherungspflichtige Tätigkeit ausgeübt werden muss. Es muss eine Arbeitslosenmeldung vorliegen. Der Antragsteller hat die Pflicht diese spätestens am 1. Tag der Beschäftigungslosigkeit abzugeben. Sind diese Punkte erfüllt, so liegen alle Voraussetzungen für einen Leistungsbezug vor.

⁵ Was als zumutbar gilt, ist dabei umstritten. Die Stelle muss jedoch für das Alter, den Gesundheitszustand und den erlernten Beruf geeignet sein. Eine ständige Abrufbereitschaft, ein Arbeitsweg von mehr als 4h täglich, Mobbing oder ein Lohn unter 70% des versicherten Verdienstes gelten ebenso als unzumutbar [19]

⁶ Länge der vorhergehenden Beschäftigung

⁷ dies gilt nur bis zum 3. Lebensjahr des Kindes

1.2.2 Dauer der Leistungen

Die Dauer des Bezuges von Leistungen hängt maßgeblich von der Dauer, des vorher ausgeübten versicherungspflichtigen Arbeitsverhältnisses ab. Die Länge des Leistungsbezuges kann dabei zwischen 6 und 24 Monaten liegen, dieser endet nur vorzeitig, wenn der Antragsteller zum Rentenempfänger wird oder in Arbeit über geht. Die maximale Bezugsdauer ist auf 12 Monate beschränkt, solange die Person jünger als 50 Jahre ist. Die Dauer der Leistungen berechnet sich in diesen Fällen, aus der Länge der vorherigen Beschäftigungen, geteilt durch 2. Ist die Person älter als 50 Jahre, so gelten gesonderte Bedingungen des Leistungsbezuges. Zusätzlich gibt die kurze Anwartschaftszeit [15]. Bei dieser muss der Arbeitnehmer mindestens 6 Monate in einem versicherungspflichtigen Arbeitsverhältnis gestanden haben. Dieses darf nicht von vorneherein auf mehr als sechs Wochen befristet sein. Das Bruttoarbeitsgeld darf jährlich nicht mehr als 30.240,00 € betragen. Die kurze Anwartschaft ist befristet bis zum 01.08.2012.

Dauer der Beschäftigung in Monaten	Art der Anwartschaft	Lebensjahr	Anspruch in Monaten
6	kurze Anwartschaft		3
8	kurze Anwartschaft		4
10	kurze Anwartschaft		5
12	Regelanwartschaft		6
16	Regelanwartschaft		8
20	Regelanwartschaft		10
24	Regelanwartschaft		12
30	Regelanwartschaft	50	15
36	Regelanwartschaft	55	18
48	Regelanwartschaft	58	24

Tabelle 1.2: Länge des Leistungsbezuges nach vorheriger Beschäftigungsdauer

1.3 SGB II Arbeitslosigkeit

1.3.1 Was ist SGB II Arbeitslosigkeit?

Zum 01.01.2005 traten die Gesetzte für Leistungen nach dem SGB II in Kraft. Diese wurden in einen 4 Stufen Plan, zum 01.01.2003 schrittweise eingeführt. Es traten Neuerungen in Kraft, bekannt und diskutiert sind vor allem die Ich-AGs, Mini Jobs und die Einrichtung von Jobcentern. Vor 2005 gab es die Sozialhilfe, die Arbeitslosenhilfe und das Arbeitslosengeld. Nach 2005 ist dieses System auf 2 Stufen geschrumpft. Es gibt Arbeitslosengeld II und Arbeitslosengeld I. Das gesetzte Ziel ist, Arbeitslose wieder in den Arbeitsalltag zu integrieren.

1.3.2 Voraussetzungen für den Bezug von SGB II

Es erhält jeder Leistung, der erwerbsfähig, hilfebedürftig⁸ und den Arbeitsmarkt zur Verfügung steht. Dies umfasst einen weiten Personenkreis.

Zu den Personen, die nicht leistungsberechtigt sind, zählen:

- Personen, welche jünger als 15 oder älter als 65 Jahre sind.
- Personen, die nichts erwerbsfähig sind.⁹
- Personen, deren gewöhnlicher Aufenthaltsort nicht in Deutschland ist.

Die Dauer des Bezuges ist an keine Grenze gebunden. Die Voraussetzungen der Arbeitslosigkeit müssen erfüllt sein. Die Leistung kann ganz oder teilweise gestrichen werden, wenn Eigenbemühungen fehlen oder andere Pflichten verletzt werden.

1.4 Übergänge und Übergangswahrscheinlichkeiten

Übergänge sind der Hauptbestandteil der Diplomarbeit. Wenn von einem Übergang gesprochen wird, so ist stets der Übergang von der SGB III Arbeitslosigkeit in die SGB II Arbeitslosigkeit gemeint. Ein Übergang wird definiert als:

- Die Personen bekommt x-Monate Arbeitslosengeld I.
- Es folgt ein Zeitraum von 1 Tag oder mehr, wobei 1 Tag die Regel ist.
- Der Antragsteller bezieht danach Arbeitslosengeld II.

Es ist zu beachten, dass Personen nur zählen, welche Leistungen nach SGB III empfangen haben. Es gibt Fälle, welche lediglich eine Beratung in Anspruch genommen haben oder sich arbeitsuchend gemeldet haben. Dieser Personenkreis wird im Folgenden nicht betrachtet. Dasselbe gilt für die SGB II-Empfänger. Des Weiteren wird definiert, dass ein gleichzeitiger Bezug von Leistungen nach dem SGB III und SGB II, als Empfang von Arbeitslosengeld II behandelt wird. Diese Personen sind von der Höhe der Leistungen auf SGB II Niveau sind. Der Zeitpunkt des Übergangs wird als letzter Tag des Arbeitslosengeldes I Bezuges festgelegt. Die Anzahl der Tage zwischen SGB III und SGB II Leistungen wird dann als direkt angesehen, wenn der Zeitraum kleiner als 8 Tage ist.

P(W)... Anteil der übergegangenen Personen

⁸ Wer seinen Lebensunterhalt nicht oder nur unzureichend aus eigenen Kräften bestreiten kann, gilt als hilfebedürftig.

⁹ Dies sind alle Personen die weniger als 3h pro Tag arbeiten können, schwerbehindert sind oder voraussichtlich länger als 6 Monate krank sein werden.

X ... Anzahl der übergegangenen Personen

Y ... Anzahl der SGB III Leistungsempfänger

$$P(W) = \frac{X}{Y} \quad (1.1)$$

2 Datenanalyse und Verarbeitung

2.1 Aufbau des Datensatzes

Den Ausgangsdatsatz bildet die IEB, den Integrierten Erwerbs-Biografien. In dieser befinden sich die Erwerbsbiografien aller Personen in Deutschland. Für die vorliegende Arbeit werden lediglich die Personendaten, welche in den Jahren 2007 bis 2009 Leistungen nach SGB II oder SGB III bezogen haben ausgewertet. Der Datensatz besteht aus 54630594 Spells¹⁰. Insgesamt sind die Daten von 857102 Personen gespeichert. Jede Person kann dabei mehrere Spells haben. Ursachen für den Beginn eines neuen Spells sind vielfältig.

- Die Person nimmt eine neue Beschäftigung auf.
- Es erfolgt eine Teilnahme an einer arbeitsmarktpolitischen Maßnahme.
- Es werden Sanktionen verhängt oder personenbezogene Merkmale ändern sich.

Jede Änderung bringt einen neuen Spell hervor. Dadurch ist es möglich, dass für den Zeitraum von 2007 bis 2009, eine Person 100 oder mehr Spells haben kann. In der Auswertung findet dies Berücksichtigung. Viele Spells überschneiden sich in den Zeiträumen, weshalb ein Episodensplitting durchgeführt wird. Ziel der Aufbereitung ist es alle Personen herauszufiltern, welche einen Übergang vom SGB III in den SGB II Bereich haben. Es werden alle Personen benötigt, welche SGB III bezogen haben. Die Gruppe dient dabei als Vergleich. Die IEB wird aus 4 Quellen zusammen gespielt. Dies sind die Beschäftigten-Historik¹¹, die Leistungsempfänger-Historik¹², die Maßnahme-Teilnehmer-Historik¹³ und die Arbeitsuchenden und Bewerberangebotsdaten¹⁴. Um die Datei Speicherplatz orientiert und übersichtlich zu gestalten, werden eine ganze Reihe von Variablen entfernt. Betroffen sind Daten zu Arbeitsverhältnissen, Zählvariablen oder Angaben zu Gründen des Zu- bzw. Abgangs aus einem Status. Folgende Tabelle gibt einen Überblick über die Variablen, welche in den folgenden Berechnungen verwendet werden.

2.2 Episodensplitting

Um die Problematik von überschneidenden Zeiträumen zu umgehen, wird ein Episodensplitting angewandt. Diese Überschneidungen entstehen, wenn mehrere Beschäfti-

¹⁰ Zeile eines Datensatzes

¹¹ enthält sozialversicherungspflichtige Beschäftigungen

¹² enthält Daten zum Empfang von Leistungen

¹³ enthält Daten zu arbeitsmarktpolitischen Maßnahmen

¹⁴ enthält Daten zur Arbeitssuche

Variablenname	Inhalt
ieb_prs_id	Personen ID, jede Person kann damit eindeutig zugeordnet werden.
ieb_geb_dat	Geburtsdatum
ieb_sex_id	Geschlecht
ieb_wo_gst_num	Geschäftsstelle am Wohnort
ieb_wo_krs_num	Kreis dem der Wohnort zugeordnet ist
ieb_sbs_id	Schwerbehindertenstatus
ieb_bsb_id	Schulabschluss
ieb_beg_epi	Beginn der Episode (aus Episodensplitting entstanden)
ieb_end_epi	Ende der Episode (aus Episodensplitting entstanden)
ieb_comb_quelle	Ist als Bitmuster angelegt, daraus kann bestimmt werden, welche Quellen einfließen
ieb_staat_num	Staatsangehörigkeit
ieb_erw_stat	Erwerbsstatus in der Episode

Tabelle 2.1: Übersicht der verwendeten Variablen

gungen oder Maßnahmen in einen Zeitraum fallen. Die Daten werden so zerlegt, dass es zwar immer noch parallele Spells gibt, diese sich aber nicht mehr vom Datum überschneiden. Durch das Splitting werden die Datensätze verdoppelt. Damit verlängert sich die Laufzeit der Programme.

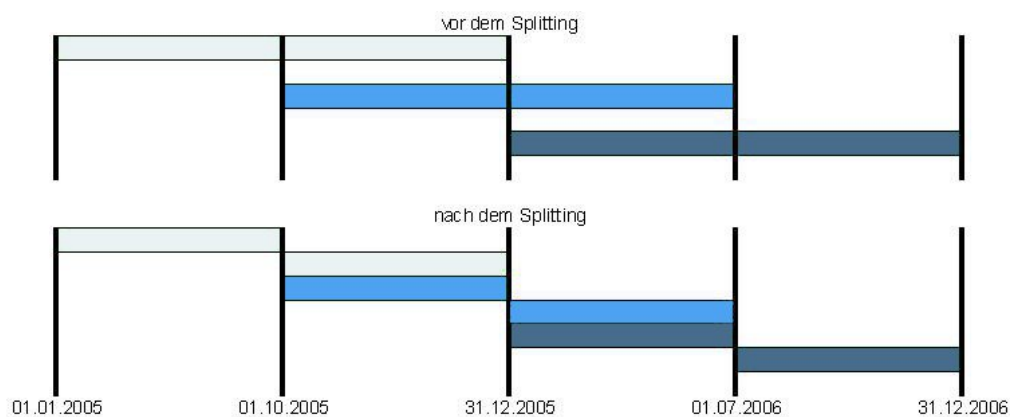


Abbildung 2.1: Anzahl der Spells vor und nach dem Episodensplitting

	Episoden ID	von	bis
ohne Splitting	1	01.01.2005	31.12.2005
	2	01.10.2005	30.06.2006
	3	01.01.2006	31.12.2006
mit Splitting	1	01.01.2005	30.09.2005
	2	01.10.2005	31.12.2005
	3	01.10.2005	31.12.2005
	4	01.01.2006	30.06.2006
	5	01.01.2006	30.06.2006
	6	01.07.2006	31.12.2006

Tabelle 2.2: Spells vor und nach dem Episodensplitting

2.3 Datenquellen

Die IEB besteht aus mehreren Quellen, welche zusammen gespielt werden. Dies geschieht über die Personen ID, welche personengebunden ist. Ein mehrfaches Auftreten ist damit nicht möglich. Einen Überblick über die wichtigsten Quellen und ihre Bedeutung, gibt der folgende Abschnitt [17].

Beschäftigten Historik (BeH)

Diese Quelle enthält die Daten, welche mit Beschäftigungsverhältnissen im Zusammenhang stehen. Betroffen sind Entgelte, Änderungsmeldungen, sowie Daten die an Renten und Sozialversicherungsträgern weiter geleitet werden. Diese Quelle fließt seit 1990 ein. Durch die langen Abfrageintervalle kommt es zu weniger dichten Daten. Wird die IEB zu Ende 2009 abgefragt, so ist damit zu rechnen, dass die Daten lediglich bis Ende 2008 vollständig sind.

Leistungsempfänger-Historik (LeH)

In dieser Quelle werden alle Bezieher von Leistungen nach dem SGB III geführt. Sie bietet damit einen Zugang zu den für die Auswertung nötigen Leitungsbeziehern.

Leistungs-Historik Grundsicherung und Leistungs-Historik-Grundsicherung XSozial(LHG und LHGX)

In den beiden Quellen werden die Empfänger von Leistungen nach den SGB II geführt. Der Grund für das bestehen zweier Quellen liegt in den unterschiedlichen Erfassungssystemen der zugelassenen kommunalen Träger (zkT) und der getrennten Trägerschaften (gT).

Maßnahmen-Teilnahme-Historik (MTH)

Diese Quelle enthält alle Maßnahmen der aktiven Arbeitsmarktförderung nach dem SGB III und SGB II. Darunter zählen unter anderen freie Förderungen, Maßnahmen der beruflichen Weiterbildung, Personal Service Agenturen, Jugendsofortprogramme, Programme für Langzeitarbeitslose, Einstiegsgelder, Berufsberatung und viele andere mehr.

Arbeitssuchend-Historik und Arbeitssuchend-Historik XSozial (ASU und XASU)

Die Personen der Quelle gelten als arbeitssuchend. Es ist nicht zwangsläufig ein Bezug von Leistungen an diese Quelle gekoppelt. Unter anderen können Personen auftreten die arbeiten jedoch befürchten demnächst ihre Beschäftigungen zu verlieren oder eine andere Beschäftigung suchen.

Für die Auswertung der Daten werden die verschiedenen Quellen benötigt die in einen Zeitraum einfließen. Die Bestimmung erfolgt über die Spalte `ieb_comb_quelle`, welche als Bitmuster angelegt ist. Für jeden Spell wird eine Bitzerlegung durchgeführt. Einen Ausschnitt der Bltmusterzuordnung enthält folgende Tabelle. Als Beispiel wird angenommen, dass der Wert von `ieb_comb_quelle` 69 ist. Für die Zerlegung werden die

Werte der einzelnen Quellen benötigt [17]: Es wird angenommen, dass der Wert von

Bitmuster	Wert	Bedeutung
2^0	1	Beschäftigten Historik (BeH)
2^1	2	Leistungsempfänger Historik (LeH)
2^2	4	Arbeitsbeschaffungsmaßnahme
2^3	8	Förderung beruflicher Weiterbildung
2^4	16	Freie Förderung
2^5	32	Europäischer Sozialfonds
2^6	64	Arbeitssuchend

Tabelle 2.3: Bitmusterzuordnung der Quellen

ieb_comb_quelle 65 ist. Der höchste auftretende Wert in 65 ist 64, dass bedeutet das die Quelle arbeitssuchend enthalten ist.

$65 - 64 = 1$ der höchste Wert in 1 ist die 1, dies entspricht der Quelle der Beschäftigten-Historik.

Auf diese Weise kann jede beliebige Kombination von Quellen erstellt werden. In den aufgezeigten Fall wäre die Person beschäftigt und gleichzeitig arbeitssuchend gemeldet.

2.3.1 Datenqualität und Vollständigkeit

Besonders bei den zugelassenen kommunalen Trägern fällt auf, dass ein Großteil Daten unvollständig sind. Davon betroffen sind der Schwerbehindertenstatus, die Staatsangehörigkeit und die Zuordnung der Kreise und Geschäftsstellen. Diese Spalten machen einen nicht unerheblichen Teil aus, weshalb die Daten nicht einfach gelöscht werden können. Die Daten werden, so weit dies möglich ist, vervollständigt und angeglichen (siehe nächstes Kapitel).

Es gab eine Systemumstellung, welche sich auf die Quelle Leistungsempfängerhistorik auswirkte. Das alte Verfahren zur Übermittlung von Daten nannte sich coLei. Das neue Verfahren nennt sich COLIBRI. Der Übergang zu dem neuen System erfolgte schrittweise, zwischen 2004 und 2006. Als 1. Agentur wurde Gotha umgestellt, nach und nach folgten die anderen Agenturen. Es musste jeder Datensatz von einem Sachbearbeiter im alten System abgemeldet und im Neuen angemeldet werden. Bis Ende 2006 waren auf diesem Weg alle Datensätze importiert wurden. Durch die verschiedenen Umstellungen gab es Probleme bei der regionalen Auswertung. Eingeführt wurde das Arbeitslosengeld II 2005 weshalb die Werte erst ab 2007 als verlässlich anzusehen sind. Problematisch sind vor allem die zugelassenen kommunalen Träger (zkT). Mit Einführung des Arbeitslosengeldes II wurde die technische Schnittstelle X-Sozial-BA-SGB II entwickelt. Die zugelassenen kommunalen Träger sind verpflichtet, monatlich Daten zu liefern. Infolge der Neuanschaffung von EDV-Verfahren dauerte es bis Mitte 2005 bis alle zugelassenen kommunalen Träger die Daten liefern konnten. Ein weiteres Problem stellten die Bedarfsgemeinschaften dar. Es mussten mehrere Datensätze zu einen ver-

eint werden. Ende 2006 war dieses Problem jedoch gelöst. Zusätzlich gab es Probleme mit der Plausibilität der Daten. Die Prüfung dauerte bis Mitte 2006 an. Betroffen waren die Bereiche der Maßnahmenteilnahme und der Meldung der Arbeitslosigkeit. Bis Anfang 2007 waren diese Probleme behoben. Weiterführende Informationen befinden sich in der Quelle [20].

2.4 Verarbeitung des Datensatzes

Beim Laden des Datensatzes werden deshalb nur die benötigten Spalten mit einbezogen. Zu Beginn war der Datensatz nach Jahren aufgeteilt, um eine schnellere Bearbeitung zu ermöglichen. Das Problem bei dieser Bearbeitung ist, dass die Übergänge die von Dezember bis Januar stattfinden nicht mehr erfasst werden. Davon betroffen sind Übergänge, welche das Ende des SGB III Anspruches im Jahr X haben und den Anfang des SGB II Anspruches im Jahre $X + 1$. Es kommt zu einem Einbruch der Übergangszahlen. Die Zuordnung der Jahre muss deshalb erst bei der Auswertung erfolgen. Die Daten vor dem 01.01.2007 werden entfernt, weil diese durch die im Kapitel 2.3.1 genannten Gründe nicht verlässlich sind. Die Auswertung erfolgt bis zum 31.12.2009. Die nachfolgenden Spells können nicht entfernt werden, da dadurch die Übergänge am Ende des Jahres 2009 verloren gehen. Die Zuordnung des Bezuges von SGB III oder SGB II Leistungen erfolgte am Anfang, über die Trägerart¹⁵ und des Erwerbsstatus. Es konnte damit lediglich eine Zuordnung des Rechtskreises erfolgen. Offen bleibt in diesen Fällen, ob die entsprechende Person wirkliche¹⁶ Leistungen erhalten hat oder lediglich arbeitsuchend gemeldet war.

Die Zuordnung erfolgt aus diesen Grund mittels Quellen. Wie oben beschrieben, gibt die Variable `ieb_comb_quelle` an, welche Quellen in den Spell zusammenfließen. Für jeden Spell wird eine Bitmusterzerlegung dieser Spalte durchgeführt. Dabei werden alle Spells gekennzeichnet in denen die 2 (LeH), 32768 (LhG) oder 262144 (XLhG) auftreten. Erhält ein Antragssteller, Leistungen nach SGB II und SGB III, so wird dies als SGB II Empfang angesehen. Der Grund hierfür ist, dass die Person dann als hilfebedürftig gilt, aber nicht mehr Leistungen bekommt als ein Bezieher von SGB II. Es liegt einzig ein Unterschied in den Trägern vor. Im nächsten Schritt erfolgt die Zuordnung der entsprechenden Leistungen. Falls bei der Quelle LeH eine Zuordnung existiert, so bezieht die entsprechende Person Leistungen nach den SGB III. In allen anderen Fällen wird zugeordnet, dass es sich um einen SGB II Bezug handelt.

In der späteren Auswertung wird untersucht, inwieweit die Ausübung von Beschäftigungen oder Maßnahmen, die Wahrscheinlichkeit einen Übergang zu haben beeinflusst. Dafür wird die Spalte `ieb_erw_stat_num` benötigt, welche den Erwerbsstatus enthält.

¹⁵ Gibt an ob der Rechtskreis SGB III oder SGB II ist. Es gibt 4 verschiedene Trägerschaften die Bundesagentur für Arbeit (BA), die Arbeitsgemeinschaften (ARGE), die zugelassenen kommunalen Träger(zkT) und die getrennten Trägerschaften (gT).

¹⁶ in Form von geldlichen Mitteln

Art der Leistung	LeH	LhG	XLhG
SGB III	1	0	0
SGB III und SGB II	1	1	0
SGB III und SGB II	1	0	1
SGB II	0	1	0
SGB II	0	0	1
SGB II	0	1	1

Tabelle 2.4: Kombinationen der einfließenden Quellen

Die Werte zwischen 10101 und 27118 sagen aus, dass die Person an einer Maßnahme teilgenommen hat. Werte zwischen 101 und 209 stehen für das Ausüben einer Nebenbeschäftigung. Die Zuordnung erfolgt als „hat Maßnahme besucht“ und „übte Nebenbeschäftigung aus“. Bei den Behindertenstatus und den Schulabschluss wird eine Fehlerbereinigung durchgeführt. In diesen Merkmalen fehlen viele Werte. Es wird angenommen, dass sich die Merkmale über die Zeit nicht ändern. Für jede Person wird ein gültiger Wert gesucht und in allen Spells der Person ersetzt. Problematisch sind Personen, welche umgezogen sind. Es wird zugeordnet, ob der Übergang in Sachsen stattfand. Besitzt eine Person einzig Spells aus einem anderen Bundesland, so wird die Person entfernt. Es gibt Leistungsbezieher, welche sehr viele Spells haben. Oftmals umfassen diese Spells einen Zeitraum von einem Tag. Die Zeiträume werden deshalb zusammengefasst. Dies geschieht nach folgenden Regeln:

- Der Zeitraum zwischen Spell x und Spell x+1 beträgt einen Tag.
- Die Leistungsart ist in Spell x und Spell x+1 gleich.
- Es darf keinen Wechsel des Status „hat Maßnahme besucht“, „übt Nebenbeschäftigung aus“, „weder noch“ geben.¹⁷

Im darauf folgenden Schritt erfolgt die Zuordnung der Übergänge. Jeder SGB III Spell wird markiert, wenn darauf ein SGB II Spell folgt. Als Übergangszeitpunkt wird das Ende des SGB III Bezuges angenommen. Es werden diejenigen Personen aussortiert, welche nie SGB III bezogen haben und jene mit Übergängen außerhalb von Sachsen. Um eine Auswertung für die Jahre zu bekommen, wird jedem Spell das entsprechende Jahr zugeordnet. Bezieht eine Person in einen Spell über 3 Jahr die gleiche Leistung, so wird die Zuordnung in allen 3 Jahren vorgenommen. Es werden einige Spalten erzeugt, welche für die spätere Auswertung relevant sind. Ein Beispiel ist die Spalte für die direkten Übergänge. Jeden Übergang wird eine 1 zugeordnet, sofern der Zeitraum zwischen SGB III und SGB II Bezug kleiner als 8 Tage ist. Des Weiteren werden Spalten für die Panelauswertung angelegt. Diese werden nur kurz vorgestellt, ausführliche Informationen folgen im Kapitel 5.

Die Ausprägungen des höchsten Abschlusses [siehe A.1] und der Schwerbehinderung

¹⁷ Dies geschah aus dem Gesichtspunkt heraus, das damit ausgeübte Maßnahme/ Beschäftigungen verloren gehen würden.

Spaltenname	Bedeutung
sgb2007	
sgb2008	markiert ob in dem jeweiligen Jahr ein SGBIII Bezug vorlag
sgb2009	
ueb2007	
ueb2008	markiert ob in dem jeweiligen Jahr ein Übergang vorlag
ueb2009	
arbeit2007	
arbeit2008	enthält die Anzahl der ausgeübten Nebenbeschäftigungen pro Jahr
arbeit2009	
abm2007	
abm2008	enthält die Anzahl der ausgeübten Maßnahmen pro Jahr
abm2009	
alter2007	
alter2008	Alter der Person im Jahr
alter2009	
ueb	Anzahl der Übergänge innerhalb der 3 Jahre

Tabelle 2.5: Spalten für Panelanalyse

werden zusammengefasst, da einige Ausprägungen zu gering besetzt sind. Der höchste Abschluss wird unterteilt in: kein Abschluss, mittlerer Abschluss¹⁸ und hoher Abschluss. Der Schwerbehindertenstatus wird in behindert und nicht behindert eingeteilt. Einige Spalten beinhalten in der Variablen des Kreises alte Gebietsstrukturen. Es erfolgte eine Umschlüsselung auf die aktuellen Gebietsstrukturen. Es wird eine Variable angelegt, welche die Hauptagenturen enthält.¹⁹

¹⁸ bis einschließlich mittlerer Reife

¹⁹ In Sachsen gibt es 10 Hauptagenturen, die Variable wurde dabei aus den Geschäftsstellen erstellt. Die Einteilung nach Geschäftsstellen war zu genau, es kam dabei zu Problem durch die geringe Besetzung der einzelnen Ausprägungen.

3 Deskriptive Analyse

3.1 Allgemeine Betrachtung des Datensatzes

Im nachfolgenden Kapitel wird eine erste Analyse der gewonnenen Werte vorgenommen. Dabei steht die Frage im Vordergrund, welchen Einfluss die Merkmale auf die Übergänge aus dem SGB III in den SGB II Bereich haben. Gibt es personenbezogene Merkmale, bei denen man annehmen kann, dass sie einen Einfluss haben? Wie verteilen sich die Übergänge auf die 3 Jahre? Gibt es Monate, in denen mehr Menschen übergehen als in anderen?

Es werden dabei Betrachtungen von 2 Standpunkten aus angestellt. Einmal werden alle Übergänge eines Jahres betrachtet. Es kann vorkommen, dass Personen mehrfach in die Berechnung einfließen. Oftmals kommt es vor, dass eine Person mehrere Übergänge pro Jahr hat. Dies führt zu Problemen bei der späteren Auswertung. Eine Betrachtung erfolgt später. Die 2. Art Übergang gibt an, ob die Person einen Übergang in dem entsprechenden Jahr hat oder nicht. Bei der logistischen Regressionsanalyse wird auf diese Übergänge Bezug genommen. Mehrere Übergänge pro Jahr führen dazu, dass die Merkmale mehr gewichtet werden, als von Personen mit einem Übergang. Das Vorgehen stellt bei Merkmalen, die nicht von der Zeit abhängig sind, kein Problem dar. Bei der Auswertung von zeitgebundenen Daten muss beachtet werden, welchen Status die Person zu einem Zeitpunkt hat.

In den Jahren von 2007 bis 2009 gibt es insgesamt 83864 Übergänge in Sachsen. Davon entfallen 29702 auf das Jahr 2007, 31100 auf 2008 und in 2009 gehen 23062 Personen über. Die wesentlich geringe Anzahl im Jahr 2009 wird deutlich. Grund hierfür ist die bereits eingangs erwähnte Datenerfassung. Spätere Betrachtungen zeigen, dass die Anzahl der Übergänge ab August/September 2009 sinkt. Der Grund hierfür ist, dass die Daten im Februar 2010 angefordert wurden. Die Agenturen haben 6 Monate Zeit zur Meldung der Daten. Die Zeit für Meldung der Arbeitsgelegenheiten beträgt 1 Jahr, womit die Daten für das Jahr 2009 nicht auswertbar sind. In der deskriptiven Analyse ist mit niedrigeren Werten in 2009 zu rechnen. Ohne das Wissen von anderen Größen besitzen die Übergangszahlen keine Aussagekraft. Es ist deshalb eine Bezugsgröße wichtig. Daher werden alle Personen gegenübergestellt, welche Leistungen nach dem SGB III empfangen haben. Es wird ein SGB III Bezug pro Jahr angenommen²⁰. Die Werte werden für die späteren Berechnungen benötigt. Folgende Tabelle gibt einen Gesamtüberblick.

Wie ersichtlich wird, gehen 2007 ca. 16,42% der Leistungsempfänger in den SGB II

²⁰ Bei einer Zählung von allen Personen mit SGB III die mehrmals pro Jahr Leistungen beziehen, würden keine aussagekräftigen Werte entstehen. Dies liegt vor allem daran, dass es sehr oft mehrere Leistungsempfänge gibt.

Jahr	Anzahl der Übergänge	Anzahl der übergegangenen Personen	Anzahl der SGB III Empfänger	Anteil der übergegangenen SGB III Empfänger in %
2007	29702	28378	172858	16,42
2008	31100	29492	190203	15,51
2009	23026	22278	183126	12,17

Tabelle 3.1: Anzahl der Übergänge und Arbeitslosengeld I Bezieher

Bereich über. Es wird deutlich dass die Werte für 2009 zu gering ausfallen. Im Mittel²¹ gehen damit 14,58% der SGB III Empfänger in den SGB II Bereich über. Im Gegensatz zum Jahr 2007 steigt die Anzahl der Übergänge und SGB III Empfänger im Jahr 2008 an. Der Anteil der übergegangenen Person nimmt dabei leicht ab.

3.2 Auswertung nach personengebundenen Merkmalen

3.2.1 Geschlecht

Ausgehend vom Geschlecht wird betrachtet, ob dieses einen Einfluss auf den Wechseln in die SGB II Arbeitslosigkeit hat. Für die Jahre 2007 bis 2009 ergeben sich folgende Werte. Im Mittel sind 60,47% der SGB III Empfänger männlich, damit ist der Anteil we-

Ausprägung	Anzahl der übergegangenen Personen			Anzahl der SGB III Empfänger			Anteil der übergegangenen SGB III Empfänger in %		
	2007	2008	2009	2007	2008	2009	2007	2008	2009
weiblich	12184	11992	7734	69517	75324	67816	17,53	15,92	11,40
männlich	16194	17500	14544	103341	114879	115310	15,67	15,23	12,61

Tabelle 3.2: Übergegangene Personen nach Geschlecht

sentlich höher als der der Frauen. Der Anteil der männlichen SGB III Empfänger liegt mit 61,03 % ebenfalls deutlich höher als bei den weiblichen Personen. Im Jahr 2007 liegt der Anteil der übergegangenen weiblichen Personen mit 17,53 % etwas höher als die der männlichen Übergänge. In den beiden anderen Jahren unterschieden sich die Anteile der übergegangenen Personen kaum voneinander. Die Anzahl der weiblichen übergegangenen Personen nimmt 2008 ab. Bei den Männlichen steigt diese an. Die Anzahl der SGB III Empfänger steigt sowohl bei den männlichen, als auch bei den weiblichen Personen an. Ausgehend von diesen Werten wird vermutet, dass das Geschlecht keinen Einfluss auf Übergangswahrscheinlichkeiten hat.

²¹ Im Folgenden wird bei Anteilen das geometrische Mittel berechnet.

3.2.2 Alter

Für das Alter ergeben sich folgende Grafiken. Es wird dargestellt, wie viele Personen in den Altern SGB III empfangen.

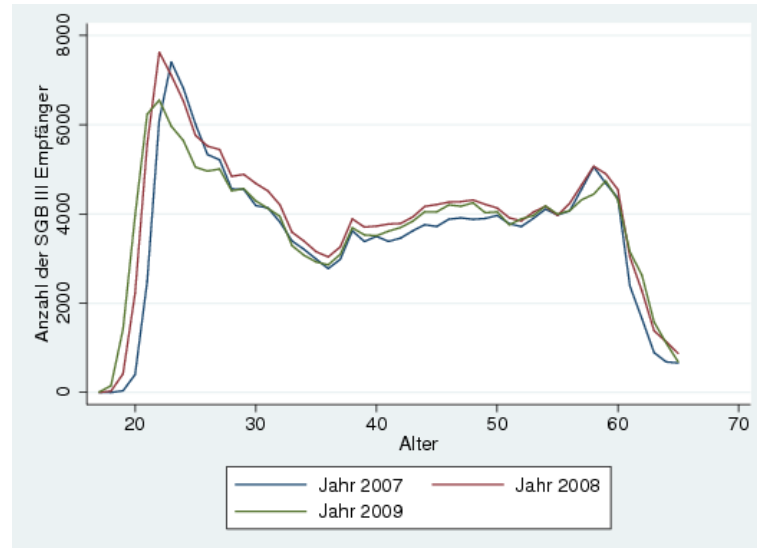


Abbildung 3.1: Verteilung der SGB III Empfänger bezogen auf das Alter

Bei der Anzahl der Übergänge ergibt sich nachfolgende Abbildung. Der Verlauf der bei-

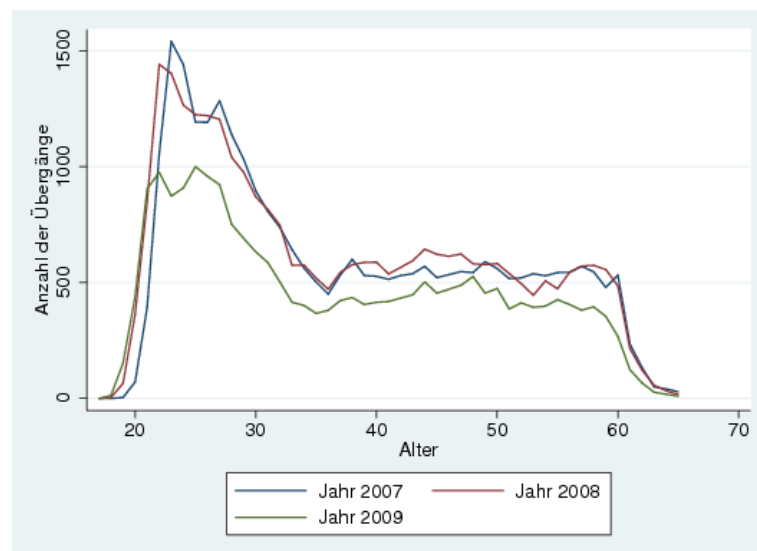


Abbildung 3.2: Verteilung der Übergänge bezogen auf das Alter

den Grafiken ist ähnlich. Bis zu einem Alter von 24 steigen sowohl die Anzahl der über-
gegangenen Personen als auch die SGB III Empfänger steil an. Bis zu einem Alter von
37 fallen die Kurven ab. Die Anzahl bleibt bei den SGB III Empfänger bis zu einem Alter
von 55 konstant. Danach steigt die Anzahl der Leistungsempfänger bis zu einem Alter

von 58 an. Die Kurve fällt bis zum Endalter 65 steil ab. Bei den übergangenen Personen ist der Verlauf bis zu einem Alter von 60 konstant. Danach folgt ebenfalls ein rascher Abstieg der Kurve. Die meisten Personen, welche Leistungen nach dem SGB III empfangen stammen aus der Altersklasse der 20 bis 24-Jährigen. Mittelt man den Anteil der 3 Jahre, so ergibt sich das 15,87% der SGB III Empfänger zwischen 20 und 24 Jahre alt sind. Die höchsten Übergangsquoten haben die 25 bis 29-Jährigen. 13,68% der Altersgruppe empfangen Leistungen nach dem SGB III, davon gehen 2007 24,48% und 2008 21,41% über. In der Gruppe der 20 bis 24-Jährigen gibt es die meisten SGB III Empfänger, dennoch gehen nur 15,61% in 2007 und 18,32% in 2008 über. Im Jahr 2007 sind 50 % aller übergegangenen Personen jünger als 35. Wie in der Grafik deutlich wird, sind die Übergänge bezogen auf das Alter rechtsschief. Dies lässt sich durch die Berechnungen der Schiefe zeigen. In allen 3 Jahren ergeben sich Werte größer 0, was auf eine deutliche Rechtsschiefheit hinweist. Im Jahr 2007 beträgt die Schiefe 0,64, 2008 0,47 und 2009 0,34. Da die Funktion nicht monoton ist, wird dies bei der Korrelation zu beachten sein.

3.2.3 Schulabschluss

Es ist anzunehmen, dass der Schulabschluss einen Einfluss auf die Anzahl der Übergänge hat. Aus Datenschutz rechtlichen Gründen werden die Kategorien des Schulabschlusses zusammengefasst. Die Kategorie des „Hauptschulabschlusses“ umfasst zusätzlich den qualifizierten Abschluss. Unter den „hohen Abschlüssen“ werden alle Abschlüsse, welche besser als die mittlere Reife sind, zusammengefasst. Der Anteil der fehlenden Werte beträgt in den Jahren weniger als 0,5%.

Ausprägung	Anzahl der übergegangenen Personen			Anzahl der SGB III Empfänger			Anteil der übergegangenen SGB III Empfänger in %		
	2007	2008	2009	2007	2008	2009	2007	2008	2009
kein Anschluss	1326	1504	1035	3996	4459	4021	33,18	33,73	25,74
Sonderschulabschluss	22	29	22	37	37	26	59,46	78,38	84,62
Hauptschulabschluss	8172	8815	6918	35435	39833	37752	23,06	22,13	18,32
mittlere Reife	16415	16716	12403	106955	115435	110535	15,35	14,48	11,22
hoher Abschluss	2339	2304	1733	25053	28760	28448	9,34	8,01	6,09

Tabelle 3.3: Übergegangene Personen nach Abschluss

Bei den Personen mit hohen Abschlüssen gehen im Mittel 7,94% in den SGB III Bereich über. 15,04% der Leistungsbezieher besitzen einen hohen Abschluss. Eine Sonderstellung nimmt der Sonderschulabschluss ein. In diesem Fall gehen im Mittel die meisten Personen über. Das ist vor allem auf die geringe Besetzung der Kategorie zurückzuführen. Für die späteren Berechnungen wird die Kategorie zu den mittleren Abschlüssen gezählt. Die meisten Übergänge treten im Bereich der Personen auf, welche keinen Abschluss aufweisen. Im Mittel gehen in dieser Ausprägung 30,65% der SGB III Empfänger in den SGBG II Bereich über. Betrachtet werden die im Mittel übergegangenen Per-

sonen. Es ergibt sich folgende Grafik: Bis auf die Kategorie der Sonderschulabschlüsse

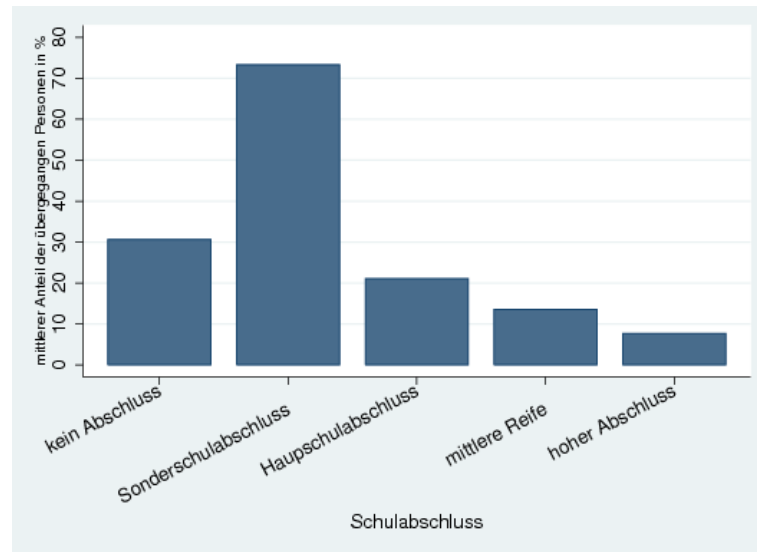


Abbildung 3.3: Im Mittel übergegangene Personen nach Schulabschlüssen

nimmt der Anteil der übergegangenen Personen mit einem höheren Abschluss ab. Die Anzahl der Übergänge und der SGB III Empfänger nimmt in 2008 zu. Lediglich die Personen mit hohen Abschlüssen haben einen leichten Rückgang in 2008. Der Anteil der übergegangenen Personen ist in allen Ausprägungen rückläufig, einzig Personen ohne Abschluss hatten mehr Übergänge.

3.2.4 Schwerbehindertenstatus

In der IEB treten folgende Ausprägungen des Merkmales Schwerbehindertenstatus auf: Anerkannt, Gleichgestellt, Gleichstellung möglich und nicht schwerbehindert. Es wird eine Zusammenfassung der Ausprägungen vorgenommen. Unterschieden wird zwischen schwerbehindert und nicht schwerbehindert. Die Personen verteilen sich wie folgt.

Ausprägung	Anzahl der übergegangenen Personen			Anzahl der SGB III Empfänger			Anteil der übergegangenen SGB III Empfänger in %		
	2007	2008	2009	2007	2008	2009	2007	2008	2009
schwerbehindert	775	704	433	4383	4529	3833	17,68	15,54	11,30
nicht schwerbehindert	27589	28786	21844	168275	185624	179277	16,40	15,51	12,18

Tabelle 3.4: Übergegangene Personen nach Schwerbehindertenstatus

Im Mittel weisen 2,33% der Personen mit Übergängen eine Schwerbehinderung auf. Der Rest gilt als nicht schwerbehindert. 2007 treten 16,4% der nicht Behinderten in den SGB III Bereich über, dem gegenüber gestellt wechseln 17,68 % der schwerbehinderten Personen. Ähnliche Zahlen ergeben sich in den beiden anderen Jahren. Im Mittel gehen

14,59% der behinderten Personen aus den SGB III in den SGB II Bereich über. Bei den nicht Schwerbehinderten sind es 14,58%. 2008 haben weniger schwerbehinderte Menschen einen Übergang als in 2007. Die Anzahl der SGB III Leistungsempfänger steigt in beiden Fällen 2008 an. Der Behindertenstatus scheint keinen Einfluss auf das Auftreten eines Überganges zu haben.

3.2.5 Staatsangehörigkeit

Nachfolgend wird erörtert, ob die Staatsangehörigkeit einen Einfluss auf das Auftreten von Übergängen hat. Die Unterscheidung erfolgt in Deutsch und Menschen mit Migrationshintergrund. Der Anteil der fehlenden Werte beträgt in diesem Merkmal 0,15%. Damit ergeben sich folgende Anzahlen und Anteile der Leistungsempfänger. 1,28% der

Ausprägung	Anzahl der übergegangenen Personen			Anzahl der SGB III Empfänger			Anteil der übergegangenen SGB III Empfänger in %		
	2007	2008	2009	2007	2008	2009	2007	2008	2009
deutsch	27833	28847	21783	170626	187466	180248	16,31	15,39	12,09
andere Staatsangehörigkeit	523	611	442	2027	2476	2501	25,80	24,68	17,67

Tabelle 3.5: Übergegangene Personen nach Staatsangehörigkeit

SGB III Leistungsempfänger sind Personen mit Migrationshintergrund. Es gehen im Mittel 22,41% dieser Personen in den SGB II Bereich über. Bei den Personen mit deutscher Staatsbürgerschaft wechseln 14,48% über. Sowohl die Anzahl der Übergänge, wie auch die Leistungsempfänger des SGB III Bereiches steigen im Jahr 2008 an. Die Auswertung lässt die Schlussfolgerung zu dass es einen Zusammenhang gibt.

3.3 Auswertung nach regionalen Merkmalen

3.3.1 Hauptagenturen

Die Variable der Hauptagentur wird aus den Geschäftsstellen erzeugt. Dies geschieht über die ersten beiden Ziffern. Nach der Auswertung ergeben sich die folgenden Werte. In den 3 Jahren gehen die meisten Menschen im Raum Leipzig über. Der Anteil der übergegangenen Personen liegt bei 17,91%. Den geringsten Anteil hat Plauen mit 12,81%. Die im Mittel übergegangenen Personen stellt folgende Grafik dar. Bis auf die Hauptagentur Bautzen steigen die Übergangszahlen im Jahr 2008 in den Hauptagenturen an. Die Anzahl der SGB III Leistungsempfänger steigt in allen Hauptagenturen stark an. Regionale Merkmale haben damit einen Einfluss auf den Anteil der übergegangenen SGB III Empfänger.

Ausprägung	Anzahl der übergegangenen Personen			Anzahl der SGB III Empfänger			Anteil der übergegangenen SGB III Empfänger in %		
	2007	2008	2009	2007	2008	2009	2007	2008	2009
Annaberg-Buchholz	2234	2287	1746	15169	16181	15406	14,73	14,13	11,33
Bautzen	4345	4288	3101	25860	27700	25554	16,80	15,48	12,14
Chemnitz	3085	3114	2425	19560	21475	20900	15,77	14,50	11,60
Dresden	3475	3609	2565	19268	21675	21551	18,04	16,65	11,90
Leipzig	5099	5495	4022	25646	28590	26761	19,88	19,22	15,03
Oschatz	2210	2253	1659	14070	15404	14300	15,71	14,63	11,60
Pirna	1715	1798	1248	11366	12395	11641	15,09	14,51	10,72
Plauen	1650	1732	1405	11732	12845	12669	14,06	13,48	11,09
Riesa	1394	1431	1051	8556	9356	8946	16,29	15,29	11,75
Zwickau	2801	2851	2345	18018	19423	19760	15,55	14,68	11,87

Tabelle 3.6: Übergegangene Personen nach Hauptagenturen

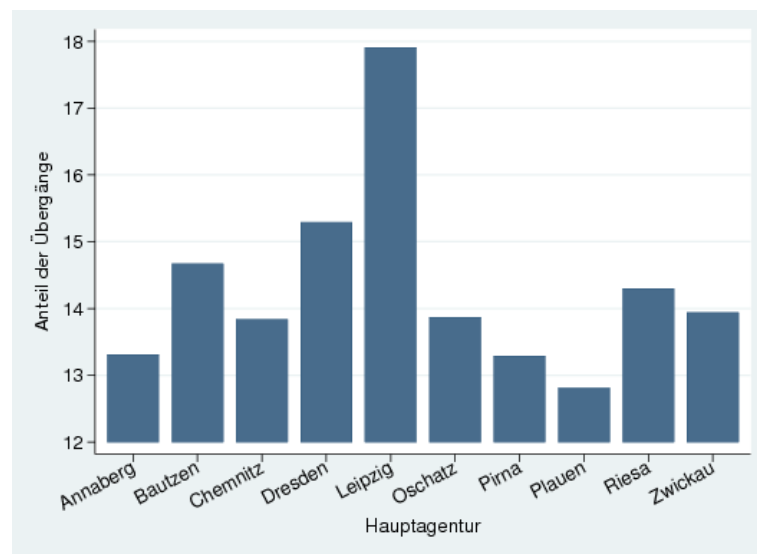


Abbildung 3.4: Im Mittel übergegangene Personen nach Hauptagenturen

3.3.2 Kreise

Die Zuordnung nach Kreisen unterscheidet sich von denen der Hauptagenturen. Verdeutlichen wird dieser Sachverhalt durch die Grafiken der Hauptagenturen ²² und der Kreise ²³. Es ergeben sich die nachfolgenden Werte für die Einteilung nach Kreisen.

Ausprägung	Anzahl der übergegangenen Personen			Anzahl der SGB III Empfänger			Anteil der übergegangenen SGB III Empfänger in %		
	2007	2008	2009	2007	2008	2009	2007	2008	2009
Stadt Chemnitz	1498	1484	1117	8190	9133	9029	18,29	16,25	12,37
Erzgebirgskreis	2760	2773	2183	18937	20282	19543	14,57	13,67	11,17
Mittelsachsen	2164	2184	1669	14406	15621	14818	15,02	13,98	11,26
Vogtlandkreis	1650	1732	1405	11731	12844	12669	14,07	13,48	11,09
Zwickau	2264	2350	1895	14160	15180	15484	15,99	15,48	12,24
Stadt Dresden	3097	3218	2256	16163	18208	17890	19,16	17,67	12,61
Bautzen	2199	2221	1719	14654	15930	15379	15,01	13,94	11,18
Görlitz	2308	2232	1513	12508	13191	11752	18,45	16,92	12,87
Meißen	1600	1649	1220	10285	11307	10942	15,56	14,58	11,15
Sächsische Schweiz- Osterzgebirge	1715	1798	1248	11366	12395	11641	15,09	14,51	10,72
Stadt Leipzig	3379	3643	2719	15304	16993	15841	22,08	21,44	17,16
Leipzig	1713	1841	1404	11288	12631	11886	15,18	14,58	11,81
Nordsachsen	1617	1674	1153	9856	10796	10063	16,41	15,51	11,46

Tabelle 3.7: Übergegangene Personen nach Kreisen

Die meisten SGB III Empfänger stammen mit 10,76% aus dem Erzgebirgskreis. Die wenigstens Arbeitslosengeld I Bezieher kommen aus Chemnitz Stadt. Der Anteil liegt bei 4,82%. Im Vergleich dazu gehen im Erzgebirgskreis 13,05% über, im Kreis Chemnitz sind es 15,43%. Leipzig Stadt weist den höchsten Wert der übergegangenen Personen auf. Es wechseln 20,10% der SGB III Empfänger in den SGB II Bereich über. Den niedrigsten Wert hat der Vogtlandkreis mit 12,81% auf. Verglichen mit den Werten der Hauptagenturen, ergeben sich ähnliche Werte. Der Kreis Leipzig Stadt ist Teil der Hauptagentur Leipzig. Die hohen Übergangszahlen in dieser Hauptagentur hängen dabei wesentlich von dem Kreis Leipzig Stadt ab. Der Kreis Leipzig weist einen geringen Anteil von 13,78% auf und liegt damit im Durchschnitt. Die Grenzen der Hauptagentur Plauen und des Vogtlandkreises sind gleich, es ergibt sich der gleiche Anteil an übergegangenen Personen. Gleiches gilt für den Kreis Sächsische Schweiz-Osterzgebirge mit der Hauptagentur Pirna.

3.3.3 Geschäftsstellen

Sachsen teilt sich in 56 Geschäftsstellen ein. Die Auswertung befindet sich im Anhang B.1. Den Ort mit dem höchsten Anteil an SGB III Empfängern im betrachteten Zeitraum

²² siehe Seite: 35

²³ siehe Seite: 35

hat Dresden mit 9,56%. Den niedrigsten Anteil an SGB III Empfänger weist Limbach Oberfrohna auf. Es kommen 0,56 % der gesamten SGB III Empfänger aus dieser Geschäftsstelle. Die wenigsten Übergänge aus den SGB III in den SGB II Bereich hat 2007 Klingenthal mit 9,34% und 2008 mit 11,04%. Den höchsten Wert hat Görlitz 2007 mit 22,92% und Leipzig 2008 mit 21,50%. Damit gehen fast ein Viertel der SGB III Empfänger später in die SGB II Arbeitslosigkeit über. Im Mittel hat Leipzig die meisten Übergänge mit 21,05%. Den geringsten Anteil besitzt Oelsnitz mit 10,37%. Es bestätigen sich die Auswertungen nach den Hauptagenturen und den Kreisen.

3.4 Auswertung nach zeitlichen Merkmalen

3.4.1 Auswertung auf Monatsebene

Betrachtet wird die Anzahl der übergegangenen Personen pro Monat im Zeitraum von Januar 2007 bis Dezember 2009:

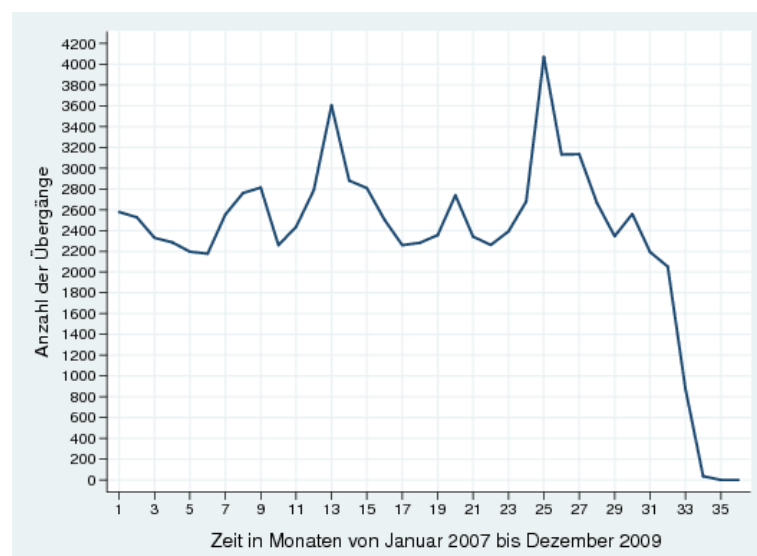


Abbildung 3.5: Hochschulabschlüsse - Anteil der Übergänge

Für die Auswertung nach Monaten wird der Fall das mehrere Übergänge pro Jahr möglich sind verwendet. Bereits ab August 2009 fallen die Werte deutlich ab. Auswertbar sind nur die Werte bis Juli 2009. Im Januar steigt die Zahl der Übergänge in jedem Jahr sprunghaft an. Einen weiteren Anstieg gibt es in den Monaten August und September. Der höchste Wert tritt im Januar 2009 mit 4070 Übergängen ein. Für eine Zeitreihenanalyse liegen zu wenige Jahre vor. Anhand der Auswertung ergibt sich einen Zusammenhang, zwischen den Anteil der übergegangenen Personen in den verschiedenen Monaten.

3.4.2 Anzahl der Übergänge pro Person

In den letzten Berechnungen wurde jeweils nur ein Übergang pro Person und Jahr betrachtet. Wie bereits erwähnt, gibt es Personen mit mehreren Übergängen. Folgende Tabelle stellt dar, wie viele Personen einen oder mehrere Übergänge besitzen. Im Mittel

Ausprägung	Anzahl der Personen			Anteil der Übergänge an der Gesamtanzahl		
	2007	2008	2009	2007	2008	2009
1	23484	23465	18040	82,75	79,56	80,98
2	4074	5049	3542	14,36	17,12	15,90
mehr als 2	820	978	696	2,89	3,32	3,12

Tabelle 3.8: Anzahl der Personen mit mehreren Übergängen

haben damit 81,09% der übergegangenen Personen einen Übergang pro Jahr. 15,75% haben 2 und 3,11 % weisen mehr als 2 pro Jahr auf. Wenngleich die Anzahl der Personen mit einem Übergang konstant geblieben ist, so wird deutlich das 2008 wesentlich häufiger 2 Übergänge pro Jahr auftreten. Die Anzahl der Personen mit mehr als 2 Übergängen ist ebenso angestiegen.

3.4.3 Zeitraum zwischen SGB III und SGB II Empfang

Ausprägung	Anzahl der Personen			Anteil der Übergänge an der Gesamtanzahl		
	2007	2008	2009	2007	2008	2009
1 Tag	19173	21121	17595	64,55	67,91	76,29
2 Tage	1170	1254	994	3,94	4,03	4,31
3 Tage	301	328	306	1,01	1,05	1,33
4 Tage	250	247	206	0,84	0,79	0,89
5 Tage	205	206	173	0,69	0,66	0,75
6 Tage	192	231	171	0,65	0,74	0,74
1 Woche	150	176	172	0,51	0,57	0,75
1-4 Wochen	2074	2323	1879	6,98	7,47	8,15
mehr als 4 Wochen	6187	5214	1566	20,83	16,77	6,79

Tabelle 3.9: Anzahl der Personen nach dem Zeitraum zwischen den SGB III und SGB II Bezug

Im Normalfall liegt rechnerisch gesehen 1 Tag zwischen dem Ende des SGB III Anspruches und den Beginn des SGB II Bezuges. Endet der Arbeitslosengeld I Bezug am 26.04, so wird ab 27.04 Arbeitslosengeld II gezahlt. Betrachtete man einen Zeitraum von 1 Woche als einen direkten Übergang, so gehen 2007 72,19% in dieser Zeit über. Im Jahr 2008 betrifft dies 75,77% und im Jahr 2009 hatten 85,06% einen direkten Übergang. Die restlichen Personen haben vor dem Bezug nach SGB II einen Zeitraum, welcher länger als eine Woche war.

3.5 Auswertung nach arbeitsmarktpolitischen Maßnahmen

3.5.1 Teilnahme an einer Maßnahme

Die Tabelle gibt darüber Aufschluss, wie viele Personen bei einem Übergang oder bei einem SGB III Bezug an einer arbeitsmarktpolitischen Maßnahme teilgenommen haben.

Ausprägung	in 2007 Übergang			in 2008 Übergang			in 2009 Übergang		
	2007	2008	2009	2007	2008	2009	2007	2008	2009
0	16614	18432	20901	21696	15951	19253	17457	15399	12247
1	7438	6800	6161	5859	8037	7977	3493	4848	6728
2 oder mehr	4326	3146	1316	1937	5504	2262	1328	2031	3303

	in 2007 SGB III Bezug			in 2008 SGB III Bezug			in 2009 SGB III Bezug		
	2007	2008	2009	2007	2008	2009	2007	2008	2009
0	122318	136848	150432	160599	130420	151652	161745	151211	130859
1	38648	25985	18700	23098	44484	30741	16267	24423	42078
2 oder mehr	11892	10025	3726	6506	15299	7810	5114	7492	10189

Tabelle 3.10: Anzahl der Maßnahmen

Personen, welche 2007 einen Übergang haben, besuchten zu 26,2% eine Maßnahme im selben Jahr. 58,4% der Personen haben im selben Jahr keine Maßnahme. Der Anteil der Maßnahmen fällt über die Jahre ab. Die Personen, welche 2007 übergehen, haben im Folgejahr mit 23,96% , 2009 mit 21,71% eine Maßnahme besucht. Fand der Übergang 2008 statt, so hatten 2007 19,87% bereits eine Maßnahme besucht. Im Jahr des Überganges sind es 27,25%. Im Folgejahr 2009 haben 27,05% der 2008 Übergegangeenen, noch einmal an einer Maßnahme teilgenommen. Betrachtet werden außerdem die SGB III Empfänger in den entsprechenden Jahren. Von den Personen die 2007 Leistungen nach den SGB III bezogen, haben 22,36% eine Maßnahme und zu 6,88% mehrere Maßnahmen besucht. Die Leistungsempfänger im Jahr 2008 hatten zu 23,39% eine und zu 8,04% mehrere Maßnahmen besucht.

3.5.2 Ausüben einer Nebenbeschäftigung

Die folgende Tabelle zeigt die Anzahl der Nebenbeschäftigungen an, die während eines Übergangs oder beim Empfang von SGB III Leistungen ausgeübt wurden. Wird davon ausgegangen, dass die Person 2007 einen Übergang hat, so haben 27,17% eine Nebenbeschäftigung. Mit einem Anteil von 29,76% gehen diese 2 oder mehr Nebenbeschäftigungen nach. 2008 haben die übergegangeenen Personen mit einem Anteil von

Ausprägung	in 2007 Übergang			in 2008 Übergang			in 2009 Übergang		
	2007	2008	2009	2007	2008	2009	2007	2008	2009
0	12224	14861	28378	17239	13765	29490	13295	13143	22277
1	7709	8755	0	8264	8017	2	6241	6002	1
2 oder mehr	8445	4762	0	3989	7710	0	2742	3133	0

	in 2007 SGB III Bezug			in 2008 SGB III Bezug			in 2009 SGB III Bezug		
	2007	2008	2009	2007	2008	2009	2007	2008	2009
0	111537	133586	172858	146436	132524	190201	146533	145740	183114
1	43787	27480	0	32267	41060	2	26977	27427	12
2 oder mehr	17534	11792	0	11500	16619	0	9616	9959	0

Tabelle 3.11: Anzahl der Nebenbeschäftigungen

30,85% eine Beschäftigung, lediglich 16,78% haben mehr als 2 Beschäftigungen. Wird angenommen, dass die Person 2008 einen Übergang hat, so gehen 28,02% davon im Vorjahr einer Nebenbeschäftigung nach. 13,53% haben 2 oder mehr Nebenbeschäftigungen. Im Jahr des Übergangs steigen diese Zahlen auf 27,16% an. 26,14% gehen dabei mehr als 2 Beschäftigungen nach. Bezogen auf die SGB III Leistungsempfänger, ergibt sich folgendes Bild. Personen mit einem Leistungsanspruch in 2007 üben mit einem Anteil von 25,33% eine Nebenbeschäftigung aus, mit 10,14% 2 oder mehr Beschäftigungen. Personen, welche 2008 Leistungen nach dem SGB III beziehen, haben mit einem Anteil von 21,59% eine Nebenbeschäftigung und mit 8,74% mehr als eine Beschäftigung. Personen aus dem SGB III Bereich gehen weniger häufiger einer Nebenbeschäftigung nach. Besonders deutlich wird dies bei der Mehrfachbeschäftigung. Aus den genannten Gründen sind die Nebenbeschäftigungen für das Jahr 2009 nicht auswertbar.

Es wird ersichtlich, weshalb 2009 in den folgenden Berechnungen ausgeschlossen wird. Die Werte sind wesentlich geringer, was darauf schlussfolgern lässt, dass diese nicht vollständig sind. Zum Anderen fehlen die Daten zu den Nebenbeschäftigungen vollständig. Diese haben jedoch einen großen Einfluss auf die Übergangswahrscheinlichkeiten.

4 Auswertung mittels logistischer Regression

4.1 Grundlagen

4.1.1 Problemstellung

Bei der logistischen Regression wird eine Auswertung durchgeführt, für eine abhängige Variable mit kategorialen Ausprägungen. Der Sachverhalt ist häufig in der Wirtschaft zu finden. Fragestellungen können dabei sein:

- Besitzt die Familie Wohneigentum? \implies dichotome Ausprägung
- Welche Blutgruppe besitzt eine Person? \implies nominale Ausprägungen
- Wie groß sind die Zukunftsängste(klein, mittel, groß)? \implies ordinale Ausprägungen

4.1.2 Zusammenhang zum linearen Regressionsmodell

Wie im linearen Modell wird bei der logistischen Regression angenommen, dass die abhängige Variable sich aus den unabhängigen Variablen zusammensetzt. Deshalb wird die logistische Regression auch als verallgemeinertes lineares Modell bezeichnet. Bei der linearen Regression wird davon ausgegangen, dass ein linearer Zusammenhang zwischen der abhängigen Variablen und den unabhängigen Variablen existiert. Es gilt dabei:

$$y_i = \alpha + \beta x_i + u_i \quad (4.1)$$

Die beiden Koeffizienten α und β werden dabei aus der Grundgesamtheit geschätzt. Dies geschieht mit der Methode der kleinsten Quadrate. Y bezeichnet man als abhängige Variable, X als die unabhängige Variable. Der stochastische Anteil wird mit u_i bezeichnet. Die lineare Regression ist nicht geeignet. Das Problem liegt in der abhängigen Variablen, die lediglich die Ausprägungen 0 oder 1 annehmen kann. Verdeutlichen lässt sich dies in einem Graphen:

Man erkennt 2 Punktwolken an den beiden Ausprägungen der abhängigen Variablen. Würde man eine Regressionsgerade durch die Punkte legen, ergibt sich folgendes Bild.

Der Zusammenhang zwischen dem Alter und der Anzahl der Übergänge ist negativ orientiert. Die abhängige Variable kann in diesen Fall Werte zwischen $+\infty$ und $-\infty$ anneh-

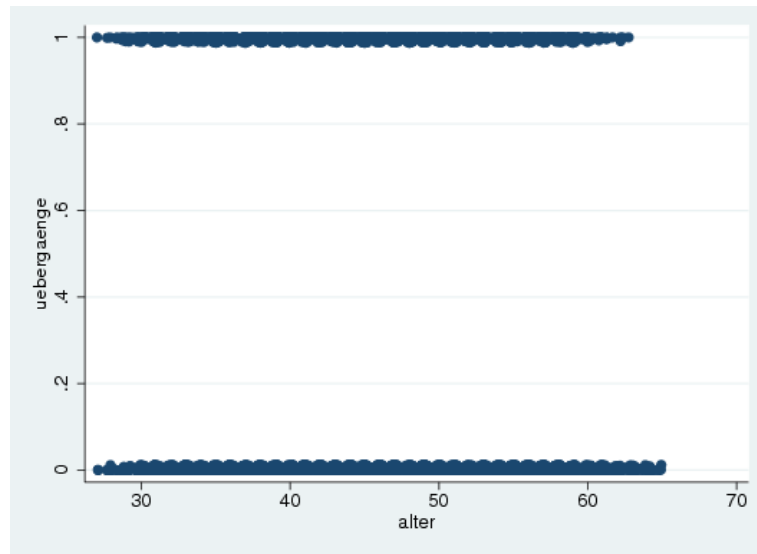


Abbildung 4.1: Alter dargestellt mit der Variablen Übergänge

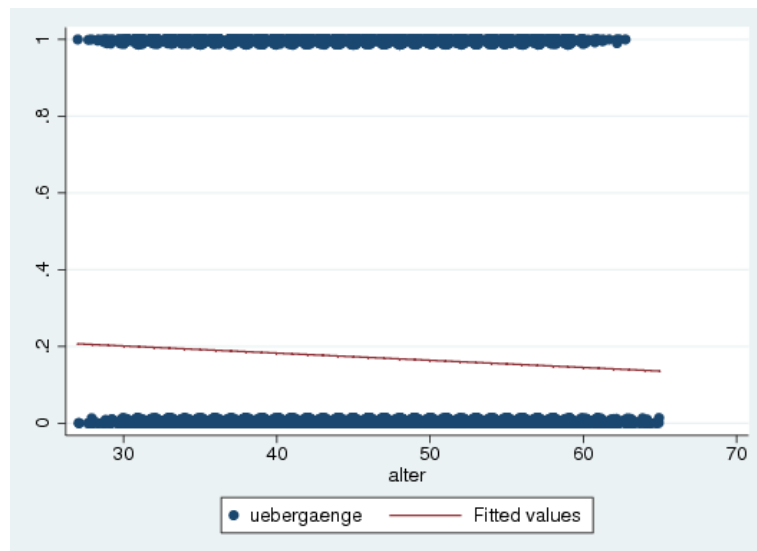


Abbildung 4.2: lineare Regressionsgerade

men. An der Grafik ist ablesbar, dass der Übergang für eine 20 Jahre alte Person ca. 0,22 beträgt. Der 20-jährige hat eine Wahrscheinlichkeit von 0,22 überzugehen. Deshalb nennt man dieses Modell lineares Wahrscheinlichkeitsmodell (LPM). Es lassen sich damit keine Werte, unter 0 oder über 1 erklären. Außerdem erkennt man das das Alter nur Wahrscheinlichkeiten zwischen 0,135 und 0,229 annehmen kann. Das stellt ein Problem dar.

Die Varianz von \hat{y} ²⁴ darf bei der linearen Regression nur konstant sein. Es wird von Homoskedastizität gesprochen. Der Scatterplot der Residuen zeigt jedoch, dass dies nicht der Fall ist. Für jeden vorhergesagten Wert gibt es 2 Residuen.

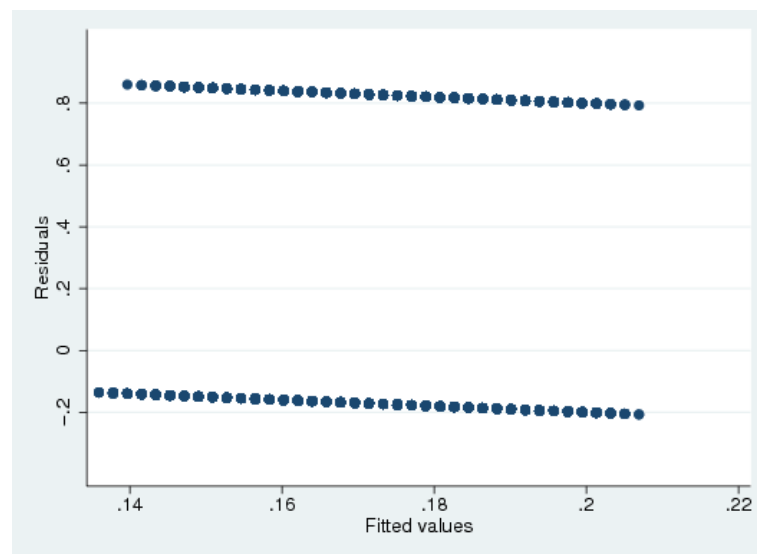


Abbildung 4.3: Verteilung der Residuen

Es gibt die Möglichkeit, dass der Übergang gleich 1 gibt, dann ist das Residuum $= 1 - \hat{y}$. Ist der Übergang $= 0$ so ist das Residuum $= 0 - \hat{y}$. Die Varianz lässt sich berechnen aus $\hat{y}_i \cdot (1 - \hat{y}_i)$. Es entstehen falsche Standardfehler, Konfidenzintervalle und Signifikanztests. Das Modell wäre damit heteroskedastisch und nicht homoskedastisch.

4.1.3 Herleitung des Logit Modells

Es wird von der einfachen linearen Regression ausgegangen. Die logistische Regressionsgleichung lautet:

$$z_k = \beta_0 + \beta_1 x_{1,k} + \beta_2 x_{2,k} + \dots + \beta_j x_{j,k} + u_k \quad (4.2)$$

²⁴ wird mit dem Befehl predict in Stata erzeugt, dieser berechnet nach jedem Regressionsbefehl die vorhergesagten Werte und schreibt diese in die Variable \hat{y}

oder

$$z_k = \beta_0 + \sum_{j=1}^J \beta_j x_{j,k} + u_k \quad (4.3)$$

Die latente²⁵ Variable z wird mittels der logistischen Funktion umgerechnet in eine Wahrscheinlichkeit p . Das Vorgehen sichert, dass die Variable y alleinig Werte zwischen 0 und 1 annehmen kann. Dafür werden die Odds berechnet.

$$Odds = \frac{p(y=1)}{1-p(y=1)} \quad (4.4)$$

Die abhängige Variable nimmt damit Wahrscheinlichkeiten zwischen 0 und $+\infty$ an.

Beispiel:

$$p(y=1) = 0,9 \Rightarrow Odds = 9$$

$$p(y=1) = 0,2 \Rightarrow Odds = 0.25$$

Der 1. Fall bedeutet, dass die Chance, dass das Ereignis $y=1$ eintritt 9-mal höher ist als, dass das Ereignis $y=0$ eintritt. Im 2. Fall tritt das Ereignis $y=0$ mit einer 4-mal höheren Chance ein.

Im nächsten Schritt wird so umgeformt das sich der Wertebereich auf $-\infty$ bis $+\infty$ ausdehnt. Es wird der \ln des obigen Ausdruckes berechnet.

$$\ln\left(\frac{p(y=1)}{1-p(y=1)}\right) = \beta_0 + \beta_1 x_{1,k} + \beta_2 x_{2,k} + \dots + \beta_j x_{j,k} + u_k \quad (4.5)$$

Es wird nach $p(y=1)$ umgeformt. Damit lässt der Ausdruck Werte kleiner 0 zu.

$$p(y=1) = \frac{1}{1 + e^{-z}}$$

$$z_k = \beta_0 + \beta_1 x_{1,k} + \beta_2 x_{2,k} + \dots + \beta_j x_{j,k} + u_k \quad (4.6)$$

Man bezeichnet diese Funktion als Linkingfunktion, da sie eine Verbindung zwischen dem Auftreten des Ereignisses $Y_k=1$ und den unabhängigen Variablen X_k herstellt. Die z_k werden als Logits bezeichnet.

Die abhängigen Werte bewegen sich zwischen 0 und 1.

Das logistische Regression Modell lässt sich beschreiben über:

$$L_k = b_0 + b_1 x_{1,k} + b_2 x_{2,k} + \dots + b_{J-1} x_{J-1,k} \quad (4.7)$$

Wird x um eine Einheit erhöht, so steigt nicht wie bei dem linearen Modell die Wahrscheinlichkeit, sondern es steigt die logarithmierte Chance.

²⁵ die Variable ist nicht direkt beobachtbar

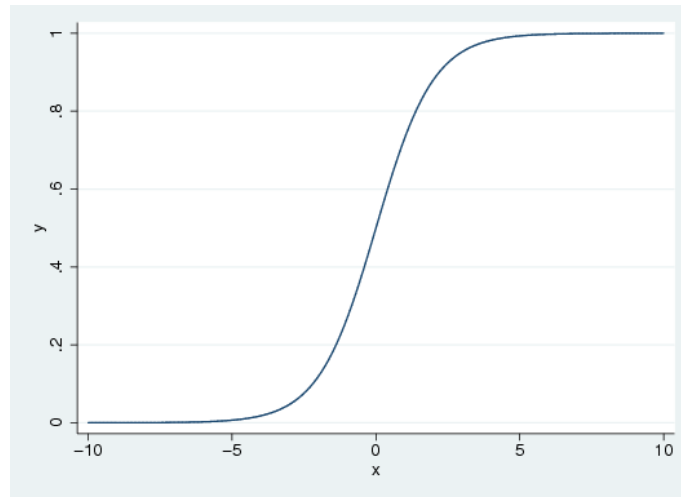


Abbildung 4.4: Verlauf der Linkingfunktion

4.1.4 Bestimmung der Koeffizienten

Bei der linearen Regression werden die Koeffizienten mittels des OLS²⁶ Verfahrens hergeleitet. Theoretisch wäre es möglich, die Koeffizienten auf diese Weise zu berechnen. Das Verfahren wäre damit äußerst ineffizient. Die Werte werden numerisch über das Maximum-Likelihood-Prinzip bestimmt. Es wird davon ausgegangen, dass die Stichprobe einen Umfang von n hat. In dieser Stichprobe tauchen h Personen auf, die das Merkmal „Übergang“ haben. Damit beträgt die Wahrscheinlichkeit:

$$P(h|\pi, n) = \binom{n}{h} \pi^h (1 - \pi)^{n-h} \quad (4.8)$$

Es soll der Koeffizienten π bestimmt werden, welcher Wert am Wahrscheinlichsten ist.

$$\mathbb{L}(h|\pi, n) = \binom{n}{h} \pi^h (1 - \pi)^{n-h} \quad (4.9)$$

Der Term $\binom{n}{h}$ wird nicht betrachtet, da dieser für alle π konstant ist. Es werden für jedes mögliche h die Likelihoods berechnet. Daraus ergibt sich für die Übergänge folgender Graph:

Abbildung 4.5: Verlauf der Likelihoodfunktion

Das Maximum liegt bei einem π von ungefähr 0,18. Das ist die maximale Likelihood-

²⁶ ordinary least squares estimation, Methode der kleinsten Quadrate

schätzung für den Anteil der Personen, die übergehen werden. Ausgehend von 4.7 werden die Koeffizienten bestimmt.

$$\pi = \hat{P}(Y = 1) = \frac{e^{b_0 + b_1 x_{1,k} + b_2 x_{2,k} + \dots + b_{J-1} x_{J-1,k}}}{1 + e^{b_0 + b_1 x_{1,k} + b_2 x_{2,k} + \dots + b_{J-1} x_{J-1,k}}} \quad (4.10)$$

Es gilt:

$$L(b_k | f, n, m) = \hat{P}(Y = 1)^h \times (1 - \hat{P}(Y = 1))^{n-h} \quad (4.11)$$

$$\left(\frac{e^{b_0 + b_1 x_{1,k} + b_2 x_{2,k} + \dots + b_{J-1} x_{J-1,k}}}{1 + e^{b_0 + b_1 x_{1,k} + b_2 x_{2,k} + \dots + b_{J-1} x_{J-1,k}}} \right)^h \times \left(1 - \frac{e^{b_0 + b_1 x_{1,k} + b_2 x_{2,k} + \dots + b_{J-1} x_{J-1,k}}}{1 + e^{b_0 + b_1 x_{1,k} + b_2 x_{2,k} + \dots + b_{J-1} x_{J-1,k}}} \right)^{n-h} \quad (4.12)$$

Theoretisch könnte von dieser Gleichung die 1. Ableitung gebildet werden und diese 0 gesetzt werden. Eine analytische Lösung zu finden ist so aber nicht möglich, deshalb werden in der Praxis iterative Verfahren eingesetzt.

Folgende Voraussetzungen müssen gelten:

- Die unabhängigen Variablen können sowohl kategorial als auch metrisch sein.
- Die abhängige Variable muss dichotom sein.
- Eine Verteilungsannahme ist nicht notwendig.
- Es werden Stichproben mit einem Umfang größer als 100 benötigt.
- Die unabhängigen Variablen müssen unkorreliert sein. Bestehende Multikollinearität²⁷ ist entsprechend auszugleichen.

²⁷ Multikollinearität tritt dann auf, wenn zwischen 2 oder mehr unabhängigen Variablen in der Regressionsanalyse, eine starke Korrelation auftritt. Als Folge dessen kann das Verfahren instabil werden und die Aussagen des Modells ungenau werden.

4.2 Analyse mittels logistischer Regression

Problematisch bleibt der Umgang mit Personen welche 3-mal in den 3 Jahren übergehen. Allgemein sieht das logistische Modell vor, das ein Datensatz in der Form: Person geht über, Person geht nicht über vorliegt.

Die logistische Regression wird in Stata mittels des Befehls *logit* oder *logistic* ausgeführt. Dabei besteht der Unterschied nur in der Darstellung, bei *logistic* werden die Odds-Ratios ausgegeben, bei *logit* die B-Koeffizienten.

4.2.1 Auswahl der unabhängigen Variablen

In diesem Kapitel soll untersucht werden welche unabhängigen Variablen Einfluss auf das Vorliegen eines Überganges haben. In der nachfolgenden Tabelle sind diese aufgeführt.

Variablenname	Beschreibung	Ausprägungen
abhängige Variable		
uebergaenge	gibt an, ob ein Übergang in den 2 Jahren stattfand oder nicht	ja nein
unabhängige Variablen		
alter	Alter bei Übergang	Alter in Jahren
ieb_sex_id	Geschlecht	männlich weiblich
ieb_staat_num	Staatsangehörigkeit	Deutsch mit Migrationshintergrund keine Angabe
schwerbe	Behindertenstatus	behindert nicht behindert keine Angabe
hoechsterab	höchster Abschluss bei Übergang	kein Abschluss mittlerer Abschluss hoher Abschluss keine Angabe
hauptagentur	zuständige Hauptagentur bei Übergang	Zuordnung siehe Anhang A.2
ieb_wo_gst_num	Geschäftsstelle bei Übergang	Zuordnung siehe Anhang A.3
ieb_wo_krs_num	Kreis bei Übergang	Zuordnung siehe Anhang A.4
abmX	Anzahl der arbeitsmarktpolitischen Maßnahmen im Jahr X	
arbeitX	Anzahl der Nebenbeschäftigungen im Jahr X	

Tabelle 4.1: Ausprägungen der verwendeten Variablen

Die Zellen sind gut besetzt, jedoch kann es trotzdem zu Problemen kommen, da die relative Anzahl sehr gering ist. Der Fakt trifft bei der Variablen des höchsten Abschlusses

zu. Es sind nur 2,25% der Absolventen ohne einen Abschluss. Bei der Staatsangehörigkeit sieht es ähnlich aus, 1,57% der SGB III Empfänger sind Menschen mit Migrationshintergrund. Bei der Behinderung kann es ebenfalls zu Problemen kommen, da der Anteil der Behinderten bei nur 2,76% liegt. Des Weiteren ist die Geschäftsstelle Brand-Erbisdorf problematisch, da diese nur bis 31.03.2008 existierte, danach wird diese mit der Geschäftsstelle Freiberg zusammengelegt.

Es wird betrachtet, inwieweit die unabhängigen Variablen Einfluss auf die abhängige Variable haben. Um die Variablen weiter zu prüfen, werden diese mit der Variablen der Übergänge korreliert:

Variable	Skalenniveau	Korrelationskoeffizient	We0e,0000rt
Hauptagentur	nominal	Cramers V	0,0507
Geschäftsstelle	nominal	Cramers V	0,0876
Kreis	nominal	Cramers V	0,0675
Behinderung	künstlich dichotom ²⁸	biseriale Korrelation	0,0431
Staatsangehörigkeit	künstlich dichotom	biseriale Korrelation	0,0297
Höchster Abschluss	ordinal	biseriale Rang Korrelation	-0,0587
Geschlecht	dichotom	Phi Koeffizient	-0,0009
Alter	metrisch	punkt biseriale Korrelation	-0,0881
arbeitsmarktpolitische Maßnahmen 2007	metrisch	punkt biseriale Korrelation	0,2234
arbeitsmarktpolitische Maßnahmen 2007	metrisch	punkt biseriale Korrelation	0,2698
Nebenbeschäftigungen 2007	metrisch	punkt biseriale Korrelation	0,3132
Nebenbeschäftigungen 2008	metrisch	punkt biseriale Korrelation	0,3448

Tabelle 4.2: Korrelation zwischen den Variablen und den Übergängen

Die Korrelation sind außer in den Bereichen der arbeitsmarktpolitischen Maßnahmen, des SGB III Bezuges und der Anzahl der Nebenbeschäftigungen, welche ausgeübt worden, sind nicht sehr hoch. Man kann davon ausgehen, dass das Geschlecht nicht mit in die Berechnung einfließen wird, ebenso wie die Staatsangehörigkeit und die Behinderung. Aus dem Bereich der Hauptagenturen, Geschäftsstellen und Kreise, kann jeweils nur eine Variable aufgenommen werden, da zwischen den Variablen eine Multikollinearität auftritt. Die Hauptagenturen werden in Geschäftsstellen eingeteilt, somit sind diese nur eine Verfeinerung der Hauptagenturen. Die Kreise sind nicht exakt mit Geschäftsstellen oder Hauptagenturen gleich zu setzen. Sie bilden Teilmengen, sodass auch auf diese Weise Multikollinearitäten auftreten. Die folgenden Karten soll dies veranschaulichen.

Aufgrund der deskriptiven Analyse wurde vermutet, dass ein Zusammenhang zwischen dem Alter und den Übergängen besteht. Nach den Korrelationen zu urteilen wird dieser wesentlich geringer als vermutet ausfallen. Das Problem besteht darin, dass der Verlauf des Alters nicht monoton ist. Für die Korrelation nach Pearson sollte diese vorliegen. Abhilfe schafft in diesen Fall das Zusammenhangsmaß Eta. Die Anzahlen der Maßnah-

²⁸ Variablen, die in Wirklichkeit nicht dichotom sind, jedoch zum Zwecke der besseren Auswertung oder Übersichtlichkeit, dichotomisiert wurden.



Abbildung 4.6: Aufteilung von Sachsen in Kreise



Abbildung 4.7: Aufteilung von Sachsen in Hauptagenturen

men und der Nebenbeschäftigungen sind monoton. Sie können ohne eine Transformation verwendet werden.

4.2.2 Kodierung der Variablen

Alle Variablen mit 2 Ausprägungen müssen in eine 0 und 1 Kodierung gebracht werden. Dabei steht 0 für einen Misserfolg und 1 für den Erfolg (oder eintreten) eines Ereignisses. Für alle Variablen, welche mehr als 2 Ausprägungen besitzen, müssen Dummy Variablen erzeugt werden. Dabei werden $k-1$ verschiedene Variablen erzeugt, welche die ursprüngliche Variable vollständig abbilden. Am Beispiel des höchsten Abschlusses ergibt sich:

hoechsterab	hoechsterab1	hoechsterab2	hoechsterab3
kein Abschluss	1	0	0
mittlerer Abschluss	0	1	0
hoher Abschluss	0	0	1
keine Angabe	0	0	0

Tabelle 4.3: Kodierung der Variablen höchster Abschluss

Durch diese Transformation wird jeder Ausprägung, eine Kodierung mit Nullen und Einsen zugeordnet. Der letzten Kategorie wird eine Kodierung aus Nullen zugeordnet. Würden alle Variablen in das Modell aufgenommen werden, so würde dies zu Multikollinearität führen. Dabei können immer nur $k-1$ Variablen hinzugefügt werden. Die verbleibende Variable nennt man Referenzkategorie. Diese Kategorie wird passend gewählt. In Stata wird automatisch die 1. erzeugte Variable als Referenz genommen, will man dies ändern, so müssen die Dummy Variablen manuell erzeugt werden. Dies geschieht durch den Befehl:

`tabulate varname, generate(new varname).`

Bei der Wahl der Referenzkategorie ist vor allem auf inhaltliche Aspekte zu achten. Bei Untersuchungen ob Abiturienten häufiger übergehen als nicht Abiturienten, so ist es sinnvoll, „hoher Abschluss“ als Referenzkategorie zu wählen. Bei ordinal skalierten Werten empfiehlt es sich, die niedrigste oder höchste Kategorie zu wählen. Außerdem muss darauf geachtet werden, dass die Kategorie ausreichend besetzt ist. Für die nachfolgenden Berechnungen werden die Referenzkategorien wie folgt festgelegt:

In Stata folgt dem Logit Befehl die abhängige Variable gefolgt von den unabhängigen Variablen. Nach Berechnung des 1. Rechenschritts ergibt sich folgendes Bild.

In diesen Schritt wird nur die Konstante in das Modell aufgenommen. Das sind alle Personen mit einem Übergang. Durch die Aufnahme von weiteren unabhängigen Variablen

Variable	Referenz
Hauptagentur	Annaberg
Geschäftsstelle	Annaberg
Kreis	Chemnitz (Stadt)
Behindertenstatus	nicht behindert
Höchster Abschluss	kein Abschluss
Geschlecht	weiblich

Tabelle 4.4: Referenzkategorien

Schritt	Regressions- koeffizient B	Standardfehler	Waldstatistik	Signifikanz	Exp(B)
0	-1,984431	0,0045421	-436,90	0,000	0,1375

Tabelle 4.5: 1. Ergebnis nach Aufnahme der Konstanten

soll eine bessere Trennung der beiden Gruppen erfolgen. Angezeigt werden die Koeffizienten, die Waldstatistik²⁹ und die Signifikanz. In Stata lässt sich mittels des Befehls *estat class* die Klassifikationstabelle für die aktuelle Berechnung aufrufen. Demnach sind nur durch die Aufnahme der Konstanten 87,92% der Fälle richtig klassifiziert. Die Sensitivität³⁰ dieser Berechnung beträgt jedoch 0. Die Spezifität³¹ beträgt 100,00%. Als Cut-Off Wert wird in Stata standardmäßig 0,5 angenommen. Das heißt, dass alle Wahrscheinlichkeiten $\geq 0,5$ als Nicht-Übergang erkannt werden. Für die Untersuchung ist es wichtig, dass die Übergänge korrekt erkannt werden. Es wird eine möglichst hohe Sensitivität angestrebt.

Als Nächstes werden die restlichen Variablen in das Modell aufgenommen. Wichtigstes Aufnahmekriterium ist dabei das χ^2 -Maß. Aufgenommen wird jene Variable mit den größten χ^2 -Wert, welche zugleich signifikant ist. Dieses Verfahren wird solange wiederholt, bis alle Variablen in das Modell aufgenommen sind oder das Modell die beste Anpassung besitzt.

Die Anzahl der Nebenbeschäftigungen im Jahr 2008 haben den größten Einfluss, auf die Wahrscheinlichkeit einen Übergang in den SGB II Bereich zu haben. Die Variable fließt als Erste in das Modell ein. Aufgrund der angestellten Betrachtungen wird die Variable Kreis als Ortsvariable aufgenommen. Der χ^2 -Wert ist im Falle der Geschäftsstellen zwar höher, es treten jedoch Korrelationen zwischen den verschiedenen Geschäftsstellen auf. Die Variable ist wegen der hohen Anzahl an Freiheitsgraden abzuwählen. Die Hauptagenturen werden damit nicht in das Modell aufgenommen, da hier der χ^2 -Wert kleiner als bei der Betrachtung auf Kreisstellenebene ist. Das Geschlecht und die Staatsangehörigkeit werden aufgrund des geringen Anpassungsmaßes ebenfalls weggelassen. Das weitere Vorgehen besteht darin nach jeder Aufnahme erneut

²⁹ Die Wald Statistik ist ein Test der Hypothese das der Standardfehler gleich 0 ist

³⁰ Unter der Sensitivität versteht man, die Fähigkeit einen realen Übergang auch als Übergang zu erkennen.

³¹ Gibt den Anteil an, dass ein Nicht-Übergang auch wirklich als Nicht-Übergang erkannt wird.

Variable	χ^2 -Wert	Freiheitsgrade
Nebenbeschäftigung 2008	38661,74	1
SGB III Bezug 2007	36489,57	1
Nebenbeschäftigung 2007	32313,08	1
Maßnahme 2008	23733,09	1
Maßnahme 2007	16396,55	1
SGB III Bezug 2008	6484,89	1
Höchster Abschluss	4350,53	3
Alter	3574,28	1
Schwerbehinderung	1806,84	2
Geschäftsstelle	1374,58	51
Kreis	827,53	13
Staatsangehörigkeit	600,40	1
Hauptagentur	410,13	10
Geschlecht	182,93	1

Tabelle 4.6: χ^2 -Werte und Freiheitsgrade nach den 1. Berechnungsschritt

die χ^2 -Maße zu berechnen und sie dementsprechend aufzunehmen. Es wird darauf geachtet, dass die Ergebnisse Signifikant sind. Zusammenfassend wird das gesamte Modell dargestellt. Dabei wird auf die Güte des Modells näher eingegangen.

Variable	χ^2 -Wert	Freiheitsgrade	richtige klassifiziert	Sensitivität	Spezifität	LogLikelihood	Cox&Snell	Nagelkerke
Konstante	0,00	0	87,92	0,00	100,00	-168173,40	0,000	0,000
Nebenbeschäftigungen 2008	38661,74	1	88,44	8,27	99,46	-148842,53	0,081	0,156
SGB III Bezug 2007	70185,89	2	88,79	15,19	98,91	-133080,46	0,143	0,273
Maßnahmen 2008	82471,96	3	88,99	21,03	98,33	-126937,42	0,165	0,317
Nebenbeschäftigungen 2007	88927,15	4	89,35	26,63	97,97	-123709,83	0,177	0,340
SGB III Bezug 2008	89070,17	5	89,33	26,57	97,96	-123638,32	0,177	0,340
Alter	92205,25	6	89,55	26,97	98,16	-122070,78	0,183	0,351
Maßnahmen 2007	94349,82	7	89,57	28,00	98,03	-120998,50	0,187	0,358
Höchster Abschluss	95566,52	10	89,63	28,38	98,05	-120390,14	0,189	0,362
Kreis	97101,52	23	89,70	29,05	98,03	-119622,64	0,192	0,368

Tabelle 4.7: Übersicht über die Entwicklung der Güte des Modells

Zum einen kann das Modell über einen ansteigenden χ^2 -Wert beurteilt werden. Da die richtig klassifizierten Fälle im Vordergrund stehen, werden diese ebenso betrachtet. Die Größe -2LL wird dabei als Devianz definiert, dabei ist -2LL das -2fache der Likelihood. Die Devianz ist χ^2 -verteilt mit $(K-J-1)$ Freiheitsgraden. K stellt den Stichprobenumfang und J die Anzahl der Parameter dar. Idealerweise sollte -2LL gleich 0 werden, dann wäre das Modell perfekt. Bei der Likelihood müsste es 1 werden, um exakt zu sein. Die Hypothesen für den -2LL Test lauten dabei:

H_0 : Modell besitzt eine perfekte Anpassung

H_1 : Modell besitzt keine perfekte Anpassung

In der Praxis werden Pseudo- R^2 -Maße zur Beurteilung der Güte eingesetzt. Betrachtet werden die Maße von Cox&Snell und Nagelkerke. Beide basieren auf der Likelihoodfunktion. Berechnet wird das Maß nach Cox&Snell dabei aus [13]:

$$R_{CS}^2 = 1 - \frac{L_0^{\frac{2}{n}}}{L_M} \quad (4.13)$$

L_0 gibt dabei den Wert der Likelihoodfunktion an, wenn lediglich die Konstante im Modell ist.

L_M ist der Wert der Likelihoodfunktion mit n aufgenommen Variablen.

Das Nagelkerke Maß normiert das Maß von Cox&Snell auf den Wertebereich 0 bis 1.

$$R_N^2 = \frac{R_{CS}^2}{1 - L_0^{\frac{2}{n}}} \quad (4.14)$$

Ähnlich wie bei der linearen Regression kann dieses Maß als Bestimmtheitsmaß verwendet werden. Es gibt dabei den Anteil der Varianz an der abhängigen Variablen an. An der Entwicklung der Gütekriterien kann man sehen, dass eine gute Anpassung durch das Modell vorgenommen wird. Vergleicht man die Pseudo R^2 , so steigen diese kontinuierlich an. Als akzeptable Werte werden bei Cox&Snell über 0,2 angesehen ab 0,4 spricht man von einer guten Anpassung. Bei Nagelkerke sieht dies genauso aus, nur geht man hier davon aus, dass Werte über 0,5 eine sehr gute Anpassung bedeuten. Betrachtet man die klassifizierten Fälle, so erkennt man dass die Sensitivität steigt. Diese ist immer noch gering. Der Grund liegt in der „unbalanced“ Datenbasis, das heißt, es liegen wesentlich weniger Erfolge als Misserfolge vor. Als Cut-Off Wert, wird in Stata 0,5 angenommen. Um die Sensitivität zu erhöhen, dürfte dieser wesentlich niedriger ausfallen. Betrachtet man den Plot der Sensitivität und Spezifität, so muss entschieden werden, was als wichtiger erachtet wird. Die Erfolge richtig zu erkennen oder die Misserfolge. Da die Erfolge hier Vorrang haben, sollte die Sensitivität möglichst hoch sein.

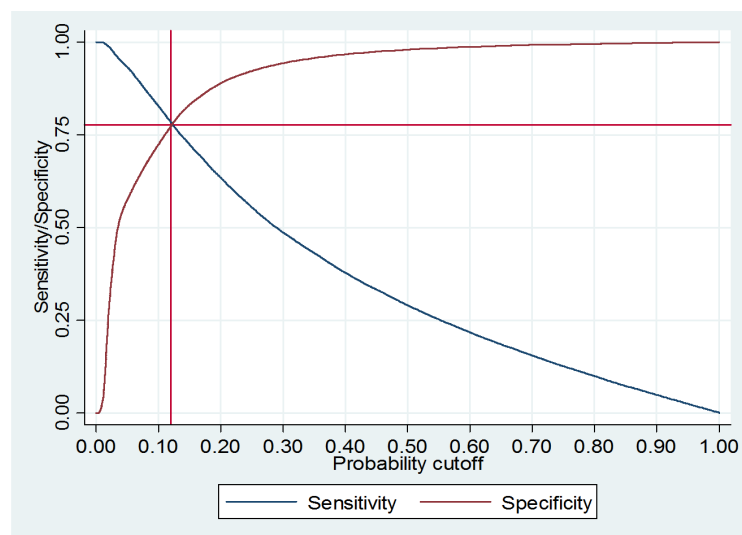


Abbildung 4.8: Sensitivität/Spezifität in Abhängigkeit des Cut-Off Wertes

Um eine gute Spezifikation in allen beiden Bereichen zu haben, sollte der Cut-Off Wert bei ca. 0,12 liegen. Damit werden alle Ereignisse, welche eine Übergangswahrscheinlichkeit unter 0,12 besitzen, als Übergang identifiziert. Alle Wahrscheinlichkeiten über

diesen Wert wären Misserfolge. Damit liegt sowohl die Sensitivität wie auch die Spezifität höher. Mittels *estat class* lässt sich die Klassifikationstabelle nach der Regression aufrufen. Bei einem Cut-Off Wert von 0,50 werden 16019 Übergänge von 55135 Gesamtübergängen erkannt. Bei den Nicht-Übergängen werden 393207 von insgesamt 401102 richtig zugeordnet. Verändert man den Cut-Off Wert auf den grafisch bestimmten Wert von 0,12, werden 47723 der realen Übergänge erkannt, im Gegensatz dazu verschlechtert sich Erkennung der Nicht-Übergänge hier wurden 298078 richtig erkannt. Damit beträgt die Sensitivität 78,98% und die Spezifität 75,19%.

Betrachtet wird die Fläche unter der Roc Kurve. Anzeigen lässt sich die Roc Kurve in Stata mittels *lroc*. Die Fläche beträgt 0,8686. Sie kann Werte zwischen 0,5 und 1 annehmen. Je höher der Wert desto besser ist die Güte des Modells und je besser ist die Trennung der beiden Gruppen. Die Fläche unter der Kurve ist dabei verteilungsunabhängig. [26]

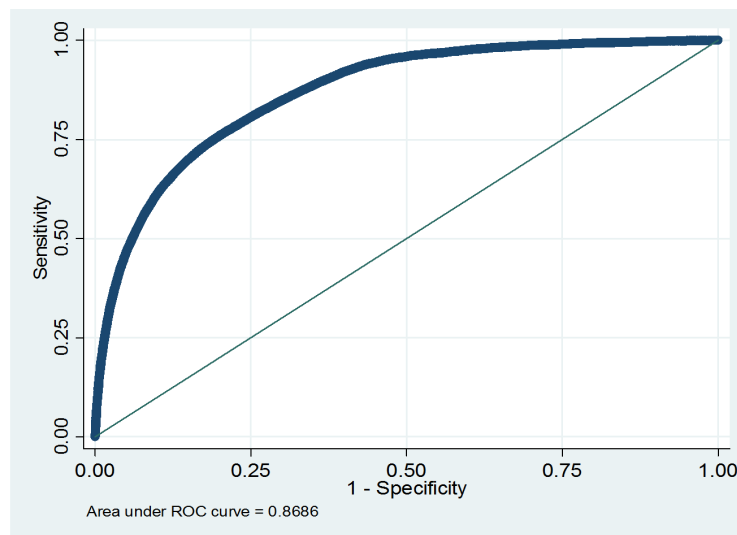


Abbildung 4.9: Roc Kurve

Es ist darauf zu achten, dass nicht zu viele Variablen aufgenommen werden. Dadurch kann ein sogenanntes Overfitting³² entstehen. Im Gegensatz dazu, gibt es das Underfitting. Es ist gekennzeichnet durch zu wenig aufgenommene Variablen. Beide Arten sind schwer zu erkennen.

4.2.3 Interpretation der Ergebnisse

Einen Überblick über das Modell gibt folgende Tabelle. Diese enthält alle aufgenommen Variablen.

³² Überanpassung eines Modells, geschieht durch das Aufnehmen von zu vielen unabhängigen Variablen. Die Schätzeigenschaften des Modells werden dadurch verringert. [24]

Variable	Regressions- koeffizient B	Standardfehler	Waldstatistik	Signifikanz	EXP(B)
Nebenbeschäftigungen 2008	0,8790509	0,0074674	117,72	0,0000	2,40861261
SGB III Bezug 2007	1,5631420	0,0122409	127,70	0,0000	4,77379698
Maßnahmen 2008	0,7760433	0,0077838	99,70	0,0000	2,17285789
Nebenbeschäftigungen 2007	0,5026406	0,0068363	73,53	0,0000	1,65308064
SGB III Bezug 2008	-0,0312597	0,0118421	-2,64	0,0080	0,96922383
Alter	-0,0222255	0,0004319	-51,46	0,0000	0,97801967
Maßnahmen 2007	0,3836039	0,0079499	48,25	0,0000	1,46756402
Mittlerer Abschluss	-0,5398098	0,0336233	-16,05	0,0000	0,58285910
Hoher Abschluss	-0,9501004	0,0357656	-26,56	0,0000	0,38670220
keine Angabe Abschluss	-1,4855940	0,0700101	-21,22	0,0000	0,22636784
Erzgebirgskreis	-0,3009229	0,0293399	-10,26	0,0000	0,74013484
Mittelsachsen	-0,2779735	0,0306444	-9,07	0,0000	0,75731689
Vogtlandkreis	-0,2613463	0,0323900	-8,07	0,0000	0,77001422
Zwickau	-0,1844022	0,0304400	-6,06	0,0000	0,83160127
Dresden	0,1375151	0,0288466	4,77	0,0000	1,14741903
Bautzen	-0,2188475	0,0302810	-7,23	0,0000	0,80344423
Görlitz	-0,0755950	0,0306675	-2,46	0,0140	0,92719164
Meißen	-0,3868038	0,0330147	-11,72	0,0000	0,67922435
Sächsische Schweiz- Osterzgebirge	-0,2784573	0,0322509	-8,63	0,0000	0,75695059
Leipzig Stadt	0,3320897	0,0285907	11,62	0,0000	1,39387787
Leipzig	-0,1106881	0,0317434	-3,49	0,0000	0,89521792
Nordsachsen	-0,0787638	0,0326671	-2,41	0,0160	0,92425821
keine Angabe Kreis	-0,4135014	0,0437882	-9,44	0,0000	0,66133061
Konstante	-2,1748550	0,0434359	-50,07	0,0000	0,11362463

Tabelle 4.8: Ergebnisse nach der vollständigen Aufnahme aller Variablen

Daraus ergibt sich die logistische Regressionsgleichung.

$Z =$

(Anzahl der Nebenbeschäftigungen 08)*0,8790509+
 (SGB III Bezug 2007)*1,5631420+
 (Anzahl der Maßnahmen 2008)*0,7760433+
 (Anzahl der Nebenbeschäftigungen 2007)*0,5026406+
 (SGB III Bezug 2008)*-0,0312597+
 (Alter in Jahren)*-0,0222255+
 (Anzahl der Maßnahmen 2007)*0,3836039+
 (Regressionskoeffizient des entsprechenden Abschlusses)+
 (Regressionskoeffizient des entsprechenden Kreises)-2,1748550

Bei der Anwendung ist der Wertebereich der Variablen zu beachten:

Der SGB III Bezug kann entweder 0 (nein) oder 1 (ja) sein. Bei den Nebenbeschäftigungen und Maßnahmen muss die jeweilige Anzahl im entsprechenden Jahr eingesetzt werden. Das Alter hat einen Bereich von 15 bis 65 Jahren. Je nach Abschluss wird der entsprechende Koeffizient eingesetzt. Im Falle, dass kein Abschluss vorliegt, muss keine Variable aufgenommen werden. Das Gleiche gilt für den Kreis. Die Konstante wird immer mit aufgenommen.

Alle aufgenommen Variablen sind signifikant. Dazu gehört ebenfalls die Wald Statistik, diese testet die Hypothese, dass der Regressionskoeffizient gleich 0 ist und damit kein Einfluss vorhanden ist. Es ist darauf zu achten die Koeffizienten nicht wie in der linearen Regression zu interpretieren. Ein negatives Vorzeichen bei den Koeffizienten deutet lediglich darauf hin, dass die Wahrscheinlichkeit sinkt, ein positives das diese steigt. Die Exponentialfunktion an der Stelle B gibt den Faktor an um den sich die Chance erhöht, wenn die unabhängige Variable um 1 erhöht wird. Je höher der Abschluss ist um so geringer ist die Chance über zu gehen, dies bezieht sich auf die Referenzkategorie „kein Abschluss“. Bei dem Alter steigt die Wahrscheinlichkeit über zu gehen, bis zu einem Alter von 27 an, danach fällt die Wahrscheinlichkeit wieder ab. Personen in den Kreisen Dresden und Leipzig Stadt haben eine höhere Chance über zu gehen als in der Referenzkategorie Chemnitz. Im Gegensatz dazu hat Meißen eine sehr viel geringe Chance. Laut der Regression müsste damit die Chance einen Übergang zu haben steigen, mit jeder Maßnahme bzw. Nebenbeschäftigung, der mehr nachgegangen wird. Die Anzahl der Maßnahmen und Nebenbeschäftigungen im Jahr 2007 wirken sich geringer aus, als im Jahr 2008. Ein SGB III Bezug im Jahr 2007 lässt die Wahrscheinlichkeit steigen einen Übergang zu haben. Ein SGB III Bezug im Jahr 2008 lässt diese fallen.

Betrachtet wird folgende Beispielperson:

Es gab keine Nebenbeschäftigungen und Maßnahmen in 2007 oder 2008. Es gibt einen SGB III Bezug in 2007. Die Person ist 25 Jahre alt, hat einen mittleren Abschluss und kommt aus dem Erzgebirgskreis.

Daraus ergibt sich folgender Wert der Regressionsgleichung:

$$z_k = b_0 + b_1x_{1,k} + b_2x_{2,k} + \dots + b_{J-1}x_{J-1,k} = -1,0080832 \quad (4.15)$$

Gemäß der Definition wird das Ergebnis mittels der Funktion $\frac{e^z}{1+e^z}$ in eine Wahrscheinlichkeit umgerechnet. Damit ergibt sich die Wahrscheinlichkeit, dass die Person übergeht, mit 0,118. Für weitere Berechnungen kann die entsprechende Excel Tabelle verwendet werden.

4.2.4 Interaktionseffekte

Bis jetzt wurde lediglich darauf eingegangen, wie sich die Hinzunahme einer Variablen auf die abhängige Variable auswirkt. Dabei hat man zwar beobachtet das Variable x_1 Einfluss auf Variable x_2 hat, dieser Einfluss war bei allen Ausprägungen gleich. Dies würde bedeuten, dass durch unterschiedliche Ausprägungen einer unabhängigen Variable, unterschiedliche Koeffizienten für eine andere unabhängige Variable auftreten können. Bei der Auswahl der Interaktionseffekte zählen vor allem die inhaltlichen Gesichtspunkte. Die Interaktionsvariable wird dabei durch einfache Multiplikation gebildet. Es kann zum Beispiel untersucht werden, ob Nebenbeschäftigungen in beiden Jahren einen Einfluss auf das Modell haben. Dafür werden die beiden Variablen arbeit2007 und arbeit2008 multipliziert. Da beide Variablen metrisch sind, würde eine 0 entstehen, wenn in einen der Jahre keine Nebenbeschäftigung ausgeübt wurde. Eine Zahl größer

als Null würde anzeigen das in beiden Jahren eine Beschäftigung vorlag. Diese Variable wird zusätzlich zu den beiden Variablen arbeit2007 und arbeit2008 aufgenommen. Mittels Stata wurden alle möglichen Interaktionen mit dem aufgestellten Modell getestet, dabei trat keine signifikante Verbesserung der Ergebnisse auf. Als inhaltlich sinnvoll wurden dabei die folgenden Kombinationen erachtet.

Variable 1	Variable 2	χ^2 -Wert	LogLikelihood	korrekt klassifiziert in %	Sensitivität in %	Spezifität in %	Pseudo R^2
Grundmodell		97101,52	-119622,64	89,70	29,05	98,03	0,2887
Nebenbe- schäftigung 2007	Nebenbe- schäftigung 2008	94607,92	-131171,91	88,37	27,23	97,69	0,2650
Maßnahme 2007	Maßnahme 2008	94613,23	-131167,25	88,37	27,29	97,68	0,2651
Nebenbe- schäftigung 2007&2008	Maßnahme 2007&2008	94936,24	-131007,75	88,37	27,49	97,65	0,2660

Tabelle 4.9: Interaktionseffekt

Die Ausprägungen der entstehenden Variablen sind dichotom. Dabei gilt das die Ausprägung 1 als ja und die Ausprägung 0 für nein steht. Keine der Variablen bringt eine Verbesserung des Modells. Es wurden des weiteren andere Variablenkombinationen getestet, auch bei diesen gibt es nur eine geringe Verbesserung bzw. eine Verschlechterung des Modells.

4.3 Spezialfall ordinale Regression

Bisher wurde davon ausgegangen, dass die Zielvariable dichotom ist. Im Folgenden wird angenommen, dass diese ordinal ist. Die abhängige Variable wird dabei erweitert auf keinen Übergang, einen Übergang und zwei Übergänge. Die Bezeichnungen beziehen sich auf die Jahre. Kein Übergang bedeutet, dass die Person weder 2007 noch 2008 einen Wechsel hatte. Ein Übergang bezeichnet die Tatsache das entweder 2007 oder 2008 ein Übergang stattfand. Im Falle zweier Übergänge findet pro Jahr einer statt. In Stata wird diese Art der Regression mittels *ologit* aufgerufen. Es folgen die abhängige Variable und die unabhängigen Variablen. Die Vorgehensweise ist dieselbe. Es soll dabei geprüft werden, ob sich eine bessere Anpassung ergibt. Da die Vereinfachung der Annahme, dass ein Übergang entweder existieren kann oder nicht mit einem Informationsverlust einhergeht.

4.3.1 Herleitung des Modells

Das Ziel der ordinalen Regression besteht darin, das Auftreten der Kategorie j einer abhängigen Variable y zu bestimmen. [12] Es wird die bedingte kumulierte Wahrscheinlichkeit berechnet, dass der Wert von y kleiner gleich j ist, unter den Einfluss der unabhängigen Variablen: $P(y/x)$. Die unabhängigen Variablen können metrisch oder kategorial sein. Kategoriale müssen in entsprechend Dummy Variablen zerlegt werden. Der Zusammenhang zwischen den tatsächlichen Übergängen und der latenten Variable y^* ergibt sich wie folgt: Y^* kann Werte zwischen $-\infty$ und $+\infty$ annehmen. Die τ geben dabei die Cut-Off Werte an [5]. Die Abstände zwischen den verschiedenen Cut-Off Werten

$y=0$	kein Übergang	falls $-\infty \leq y^* < \tau_1$
$y=1$	ein Übergang	falls $\tau_1 \leq y^* < \tau_2$
$y=2$	zwei Übergänge	falls $\tau_2 \leq y^* < +\infty$

Tabelle 4.10: Zusammenhang zwischen y und y^*

können dabei verschieden groß sein. Allgemein lässt sich folgende Gleichung aufstellen:

$$y_i = j \text{ wenn } \tau_{j-1}^* < \tau_j \quad (4.16)$$

Dabei muss gelten:

$$j=1 \dots J$$

$$\tau_0 = -\infty \quad \tau_J = +\infty$$

Die abhängige Variable ist dabei linear abhängig von den unabhängigen Variablen.

$$y_i^* = \beta_0 + \beta_1 x_{1,i} + \dots + \beta_l x_{l,i} + \varepsilon_i \quad (4.17)$$

Die zufällige Fehlervariable ist dabei ε . Betrachtet man ein einfaches Modell mit der unabhängigen Variable Abm2007^{33} in Abhängigkeit zu der Anzahl der Jahre mit Übergängen, so kommt man zu folgender Darstellung. Betrachtet wird die Wahrscheinlichkeit

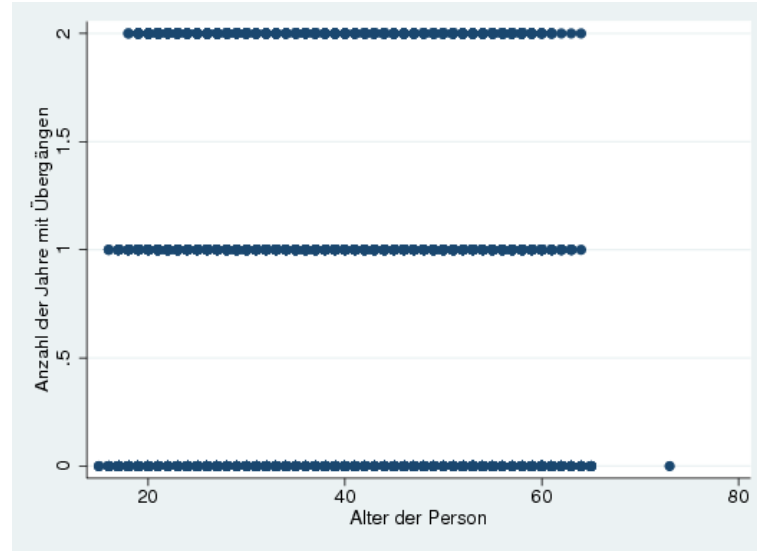


Abbildung 4.10: Anzahl der Jahre mit Übergängen - Anzahl der Maßnahmen 2007

$P(y \leq 1 | x = 1)$. Diese gibt an, dass eine Person einen oder keinen Übergang hat, unter der Bedingung dass sie an einer Maßnahme 2007 teilgenommen hat. Bei einer ordinalen logistischen Regression wird unterstellt, dass der Fehlerterm ε einer logistischen Verteilung entspricht. Die Dichtefunktion lautet:

$$f_{LOG}(\varepsilon) = \frac{e^{\varepsilon}}{(1 + e^{\varepsilon})^2} \quad (4.18)$$

Dabei ist der Mittelwert gleich 0 und die Varianz $\sigma^2 = \frac{\pi^2}{3}$. Daraus folgt die Verteilung:

$$F_{LOG}(\varepsilon) = \int_{-\infty}^{\varepsilon} \frac{e^t}{(1 + e^t)^2} dt = \frac{e^{\varepsilon}}{1 + e^{\varepsilon}} = \frac{1}{1 + e^{-\varepsilon}} \quad (4.19)$$

Ist die Verteilung der Fehlervariablen bekannt, so können die Wahrscheinlichkeiten für das Auftreten von y bei bekannten Werten der Einflussvariablen ermittelt werden.

$$P(y_i \leq 1 | x) = P(y_i^* \leq \tau_1 | x) \quad (4.20)$$

Setzt man in die Gleichung (4.17) ein, so ergibt sich.

$$P(y_i \leq 1 | x) = P(\beta_0 + \beta_1 x_{1,i} + \dots + \beta_l x_{l,i} + \varepsilon_i \leq \tau_1) \quad (4.21)$$

³³ Anzahl der Maßnahmen 2007

Der Term $\beta_0 + \beta_1 x_{1,i} + \dots + \beta_l x_{l,i}$ wird subtrahiert.

$$P(y_i \leq 1|x) = P(\varepsilon_i \leq \tau_1 - (\beta_0 + \beta_1 x_{1,i} + \dots + \beta_l x_{l,i})) \quad (4.22)$$

In die in (4.19) erstellte Verteilungsfunktion, wird der Wert $\beta_0 + \beta_1 x_{1,i} + \dots + \beta_l x_{l,i}$ eingesetzt.

$$P(y_i \leq 1|x) = F(\tau_1 - (\beta_0 + \beta_1 x_{1,i} + \dots + \beta_l x_{l,i})) \quad (4.23)$$

Verallgemeinert man diese Gleichung auf alle Kategorien $j=1$ bis J , so ergibt sich:

$$P(y_i \leq j|x) = F(\tau_j - (\beta_0 + \beta_1 x_{1,i} + \dots + \beta_l x_{l,i})). \quad (4.24)$$

Verdeutlichen soll dies das folgende Beispiel. In Stata wird die Konstante $\beta_0 = 0$ gesetzt. Das lässt sich damit begründen, dass angenommen wird, dass das Intervall für die 1. Kategorie, für welche kein Übergang vorliegt, von 0 bis τ_1 ³⁴ geht. Die Spanne für einen Übergang erstreckt sich von τ_1 bis τ_2 . Alle Werte größer als τ_2 werden der Kategorie „2 Übergänge“ zugeordnet. Als Beispiel wird das Modell betrachtet, in dem lediglich das Alter aufgenommen wird. Damit ergeben sich folgende Werte:

Regressionskoeffizient für das Alter β_1 : -0,0207813

τ_1 : 1,212203

τ_2 : 4,346687

Alter x_1 : 25

$$P(y_i \leq j|x) = F(\tau_j - \beta_1 x_1) \quad (4.25)$$

Damit ergeben sich folgende Werte, welche in die Verteilungsfunktion (4.19) eingesetzt werden:

$$\begin{aligned} P(y_i \leq 1|x = 25) &= P(\varepsilon \leq 1,7317355) = 0,849634275 \\ P(y_i \leq 2|x = 25) &= P(\varepsilon \leq 4,8662195) = 0,992356445 \end{aligned} \quad (4.26)$$

Im vorliegenden Beispiel interessieren die Wahrscheinlichkeiten, dass genau eine bestimmte Kategorie auftritt. Die Bestimmung geschieht mittels Differenzenbildung. Allgemein gilt:

$$P(y_i = j|x) = F(\tau_j - (\beta_0 + \beta_1 x_{1,i} + \dots + \beta_l x_{l,i})) - F(\tau_{j-1} - (\beta_0 + \beta_1 x_{1,i} + \dots + \beta_l x_{l,i})) \quad (4.27)$$

Für die 1. Kategorie fällt der 2. Term weg, da angenommen wird das $\tau_0 = -\infty$ und somit gleich 0 wird. Ebenso gilt für die letzte Kategorie, aufgrund von $\tau_j = +\infty$, dass in diesen Fall der 1. Term gleich 1 wird. Das obere Beispiel wird fortgesetzt. Die Wahrscheinlichkeiten für die verschiedenen Kategorien ergeben sich wie folgt:

$$\begin{aligned} P(y_i = 1|x = 25) &= 0,849634275 \\ P(y_i = 2|x = 25) &= 0,992356445 - 0,849634275 = 0,142722169 \\ P(y_i = 3|x = 25) &= 1 - 0,992356445 = 0,007643555 \end{aligned} \quad (4.28)$$

³⁴ In Stata wird die Bezeichnung cut1 verwendet.

Damit hat in dem Modell, wo nur das Alter aufgenommen wird, eine 25-Jährige Person, eine Wahrscheinlichkeit von 0,85 nicht überzugehen. Mit einer Wahrscheinlichkeit von 0,14 hat die Person einen Übergang und mit einer Wahrscheinlichkeit von 0,01 einen Übergang in beiden Jahren.

4.3.2 Probleme bei der ordinalen Regression

Identifikationsproblem

Da die latente Variable nicht messbar ist, können auch der Mittelwert und die Varianz nicht geschätzt werden. Für die Berechnung ist jedoch das Wissen über den Fehlerterm notwendig. Deshalb wurden im vorherigen Kapitel Annahmen über die Parameter getroffen. Es wurde angenommen, dass der Mittelwert gleich 0 ist und die Varianz gleich $\frac{\pi^2}{3}$ ist. Das Problem liegt in der Varianz. Nach Änderung der Varianz, werden die neuen Trennwerte τ_j^* ermittelt, so dass die Wahrscheinlichkeiten für das Auftreten des Ereignisses y gleich sind.

Warum wird $\beta_0 = 0$ angenommen? Ausgegangen wird von der Gleichung $y_i^* = \beta_0 + \beta_1 x_{1,i} + \dots + \beta_l x_{l,i} + \varepsilon_i$. Es wird $\beta_0 = \beta_0^*$ gesetzt, wobei $\beta_0^* = \beta_0 - \delta$ und $\tau_j^* = \tau_j - \delta$ ist. Es werden die Wahrscheinlichkeiten berechnet.

$$\begin{aligned} P(y_i \leq j/x) &= F(\tau_j^* - (\beta_0^* + \beta_1 x_{1,i} + \dots + \beta_l x_{l,i})) = \\ &= F((\tau_j - \delta) - ((\beta_0 - \delta) + \beta_1 x_{1,i} + \dots + \beta_l x_{l,i})) \end{aligned} \quad (4.29)$$

Damit ergibt sich, dass verschiedene Kombination von β_0 und τ_j zu gleichen Wahrscheinlichkeiten führen. In den verschiedenen Statistikprogrammen wird deshalb angenommen, dass eine dieser Variablen 0 gesetzt wird. Der andere Parameter wird dabei geschätzt. In Stata wird davon ausgegangen dass gilt: $\beta_0 = 0$. Egal was angenommen wird, die Wahrscheinlichkeiten für y bleiben unverändert.

Parallele Modellgleichungen

Es wird angenommen, dass die Regressionskoeffizienten für alle Kategorien j gleich sind. Die Steigungen für die Kurven der verschiedenen Kategorien sind lediglich verschoben. Die Verschiebung kommt durch die verschiedenen τ Werte zustande. Ist diese verletzt, so ist das Modell nicht mehr gültig. Eine Verbesserung kann durch das Zusammenfassen von Kategorien erfolgen oder durch das Ausführen einer multinomialen logistischen Regression.

4.3.3 Anwendung auf das Modell

Wie im vorherigen Kapitel beschrieben, werden die Variablen nacheinander in das Modell aufgenommen. Nach der Aufnahme der Konstanten ergibt sich damit folgendes Modell: Bei der ordinalen logistischen Regression gibt es keine Konstante. Statt dessen werden von Stata 2 Cut Werte ausgegeben. Diese entsprechen den τ aus den vorhergehenden Berechnungen. Es werden in den folgenden Schritten, die gleichen Variablen

Schritt	Standardkoeffizienten	Standardfehler	Waldstatistik	Signifikanz	Exp(B)
1,984421	0,004542	0,000000	0	7,274826762	
cut2	5,110870	0,019179	0	0	165,814219

Tabelle 4.11: 1. Ergebnis der ordinalen Regression nach Aufnahme der Konstanten

aufgenommen, welche bei der logistischen Regression berücksichtigt wurden. Die Tabelle der χ^2 -Werte ergibt sich dabei wie folgt. Die Höhe der Werte ist vergleichbar mit denen der logistischen Regression. Es gibt keine Unterschiede in der Reihenfolge. Für das Modell werden die gleichen Variablen wie in der logistischen Regression verwendet.

Variable	χ^2 -Wert	Freiheitsgrade
Nebenbeschäftigung in 2008	38448,12	1
SGB III Bezug 2007	36771,13	1
Nebenbeschäftigung in 2007	32278,01	1
Maßnahme in 2008	23582,72	1
Maßnahme in 2007	16298,95	1
SGB III Bezug 2008	6692,32	1
höchster Abschluss	4360,35	3
Alter	3568,85	1
Schwerbehinderung	1806,50	2
Geschäftsstelle	1374,84	51
Kreis	827,41	13
Hauptagentur	410,93	10
Geschlecht	187,49	1

Tabelle 4.12: χ^2 -Werte und Freiheitsgrade nach den 1. Berechnungsschritt der ordinalen Regression

Es wird der Verlauf der Gütekriterien untersucht. Anders als bei der logistischen Regression ist es nicht möglich, die richtig klassifizierten Fälle anzugeben.

Variable	χ^2 -Wert	Freiheitsgrade	Loglikelihood	Cox&Snell	Nagelkerke
Nebenbeschäftigung in 2008	38448,12	1	-159830,35	0,081	0,149
SGB III Bezug 2007	70799,02	2	-143654,90	0,144	0,264
Maßnahme 2008	82681,35	3	-137713,73	0,166	0,305
Nebenbeschäftigung in 2007	89430,63	4	-134339,09	0,178	0,327
Alter	92629,21	5	-132739,81	0,184	0,338
höchster Abschluss	93907,67	8	-132100,57	0,186	0,342
Maßnahme 2007	95947,76	9	-131080,53	0,190	0,349
Kreis	97428,73	22	-130340,00	0,192	0,354
SGB III Bezug 2008	97499,33	23	-130304,74	0,192	0,354

Tabelle 4.13: Übersicht über die Entwicklung der Güte des Modells beim ologit Modell

In diesem Modell ergibt sich eine gute Anpassung. Vergleicht man die Pseudo- R^2 -Werte der beiden Modelle, so wird ersichtlich, dass es nur geringe Unterschiede gibt. Die gegenüber Stellung der beiden Regressionen ergibt das folgende Gesamtbild. Es ist zu

beachten, dass die Werte aus den Modellen genommen werden, nachdem alle Variablen aufgenommen wurden.

Modell	χ^2 -Wert	Freiheitsgrade	LogLikelihood	Cox&Snell	Nagelkerke
logit	97101,52	23	-119622,64	0,192	0,368
ologit	97499,33	23	-130304,74	0,192	0,354

Tabelle 4.14: Gegenüberstellung der beiden Modelle

Die Modelle sind von ihrer Güte her vergleichbar. Die Loglikelihood ist beim logistischen Modell besser. Der χ^2 -Wert ist geringfügig größer im ordinalen logistischen Modell. Das logistische Modell liefert dabei leicht bessere Ergebnisse. Es kommt darauf an, in welcher Hinsicht die Ergebnisse benötigt werden. Will man die Kenntnis von Übergängen in beiden Jahren so ist das ordinale Modell besser geeignet. Wenn die Frage beantwortet werden soll, ob ein Übergang vorliegt, dann ist das logistische Modell geeigneter. Berechnet werden die Regressionskoeffizienten, um die Modellgleichung aufstellen zu können.

Variable	Regressionskoeffizient B	Standardfehler	Waldstatistik	Signifikanz	EXP(B)
Nebenbeschäftigung 2008	0,8162381	0,0068148	119,77	0,0000	2,26197449
SGB III Bezug 2007	1,5659730	0,0120345	130,12	0,0000	4,78733074
Maßnahme 2008	0,7097643	0,0072328	98,13	0,0000	2,03351190
Nebenbeschäftigung 2007	0,4966884	0,0064470	77,04	0,0000	1,64327040
Alter	-0,0221364	0,0004246	-52,13	0,0000	0,97810681
mittlerer Abschluss	-0,5401361	0,0327610	-16,49	0,0000	0,58266895
hoher Abschluss	-0,9444767	0,0349080	-27,06	0,0000	0,38888302
keine Angabe Abschluss	-1,4726680	0,0692720	-21,26	0,0000	0,22931286
abm2007	0,3678384	0,0076465	48,11	0,0000	1,44460857
Erzgebirgskreis	-0,2833030	0,0287594	-9,85	0,0000	0,75329151
Mittelsachsen	-0,2683071	0,0300505	-8,93	0,0000	0,76467291
Vogtlandkreis	-0,2615796	0,0317570	-8,24	0,0000	0,76983459
Zwickau	-0,1795932	0,0298261	-6,02	0,0000	0,83561007
Dresden	0,1310324	0,0282654	4,64	0,0000	1,14000472
Bautzen	-0,2104859	0,0297194	-7,08	0,0000	0,81019048
Görlitz	-0,0696050	0,0300669	-2,32	0,0210	0,93276219
Meißen	-0,3771893	0,0323421	-11,66	0,0000	0,68578624
Sächsische Schweiz-Osterzgebirge	-0,2691802	0,0315999	-8,52	0,0000	0,76400557
Leipzig Stadt	0,3236787	0,0279957	11,56	0,0000	1,38220313
Leipzig	-0,1069854	0,0311738	-3,43	0,0010	0,89853879
Nordsachsen	-0,0690027	0,0320076	-2,16	0,0310	0,93332416
keine Angabe Kreis	-0,4189828	0,0431093	-9,72	0,0000	0,65771551
SGB III Bezug 2008	0,0964234	0,0114701	8,41	0,0000	1,10122522
cut1	2,1924530	0,0424842			8,95715810
cut2	6,1043930	0,0476390			447,82073180

Tabelle 4.15: Ergebnisse nach der vollständigen Aufnahme aller Variablen im ologit Modell

Betrachtet wird das Beispiel aus der logistischen Regression. Es wird angenommen, dass die Beispielperson folgende Merkmale aufweist:

Es gab keine Nebenbeschäftigungen und Maßnahmen in 2007 oder 2008.

Es gab einen SGB III Bezug in 2007.

Die Person ist 25 Jahre alt, hat einen mittleren Abschluss und kommt aus dem Erzgebirgskreis.

Es ergeben sich, die folgenden Übergangswahrscheinlichkeiten:

Wahrscheinlichkeit keinen Übergang zu haben: 0,881

Wahrscheinlichkeit einen Übergang zu haben: 0,116

Wahrscheinlichkeit in beiden Jahren einen Übergang zu haben: 0,003

Bei der logistischen Regression gibt es eine Übergangswahrscheinlichkeit von 0,118.

Damit unterscheiden sich die Werte kaum voneinander. Geht man davon aus, dass durch die logistische Regression ein Informationsverlust einhergeht, so ist die ordinale logistische Regression besser als die logistische Regression trotz einer etwas schlechteren Anpassung. Es werden im Folgenden die verschiedenen Modelle der beiden Regressionen verglichen.

Variable	Regressions- koeffizient B		Standardfehler		Waldstatistik		EXP(B)	
	logit	ologit	logit	ologit	logit	ologit	logit	ologit
Nebenbeschäftigungen 2008	0,8791	0,8162	0,0075	0,0068	117,72	119,77	2,4086	2,2620
SGB III Bezug 2007	1,5631	1,5660	0,0122	0,0120	127,70	130,12	4,7738	4,7873
Maßnahmen 2008	0,7760	0,7098	0,0078	0,0072	99,70	98,13	2,1729	2,0335
Nebenbeschäftigungen 2007	0,5026	0,4967	0,0068	0,0064	73,53	77,04	1,6531	1,6433
SGB III Bezug 2008	-0,0313	0,0964	0,0118	0,0115	-2,64	8,41	0,9692	1,1012
Alter	-0,0222	-0,0221	0,0004	0,0004	-51,46	-52,13	0,9780	0,9781
Maßnahmen 2007	0,3836	0,3678	0,0079	0,0076	48,25	48,11	1,4676	1,4446
mittlerer Abschluss	-0,5398	-0,5401	0,0336	0,0328	-16,05	-16,49	0,5829	0,5827
hoher Abschluss	-0,9501	-0,9445	0,0358	0,0349	-26,56	-27,06	0,3867	0,3889
keine Angabe Abschluss	-1,4856	-1,4727	0,0700	0,0693	-21,22	-21,26	0,2264	0,2293
Erzgebirgskreis	-0,3009	-0,2833	0,0293	0,0288	-10,26	-9,85	0,7401	0,7533
Mittelsachsen	-0,2780	-0,2683	0,0306	0,0301	-9,07	-8,93	0,7573	0,7647
Vogtlandkreis	-0,2613	-0,2616	0,0324	0,0318	-8,07	-8,24	0,7700	0,7698
Zwickau	-0,1844	-0,1796	0,0304	0,0298	-6,06	-6,02	0,8316	0,8356
Dresden	0,1375	0,1310	0,0288	0,0283	4,77	4,64	1,1474	1,1400
Bautzen	-0,2188	-0,2105	0,0303	0,0297	-7,23	-7,08	0,8034	0,8102
Görlitz	-0,0756	-0,0696	0,0307	0,0301	-2,46	-2,32	0,9272	0,9328
Meißen	-0,3868	-0,3772	0,0330	0,0323	-11,72	-11,66	0,6792	0,6858
Sächsischer Schweiz	-0,2785	-0,2692	0,0323	0,0316	-8,63	-8,52	0,7570	0,7640
Osterzgebirge								
Leipzig Stadt	0,3321	0,3237	0,0286	0,0280	11,62	11,56	1,3939	1,3822
Leipzig	-0,1107	-0,1070	0,0317	0,0312	-3,49	-3,43	0,8952	0,8985
Nordsachsen	-0,0788	-0,0690	0,0327	0,0320	-2,41	-2,16	0,9243	0,9333
keine Angabe Kreis	-0,4135	-0,4190	0,0438	0,0431	-9,44	-9,72	0,6613	0,6577
Konstante	-2,1749		0,0434		-50,07		0,1136	
cut1		2,1925		0,0425				8,9572
cut2		6,1044		0,0476				447,8207

Tabelle 4.16: Vergleich der Modelle nach der ologit und logit Regression

Erkennen kann man Veränderungen vor allem bei den Nebenbeschäftigungen 2008, den Maßnahmen 2008 sowie den SGB III Bezug im Jahr 2008. Die anderen Variablen unterscheiden sich kaum von der logistischen Regression. Einen großen Einfluss auf

die Übergangswahrscheinlichkeiten haben zudem die Cut-Off Werte, diese sind vergleichbar mit der Konstanten in der logistischen Regression. Die Konstante ist negativ orientiert, das heißt, dass die Wahrscheinlichkeit abnimmt. Die Cut-Off Werte sind positiv orientiert.

Zusammenfassend lassen sich folgende Beziehungen am Beispiel des Logit Modells aufstellen. Analog sind die ologit Ergebnisse zu interpretieren. Mit jeder Maßnahme und Nebenbeschäftigung steigt die Chance eines Überganges. Bei der Nebenbeschäftigung 2008 steigt die Chance mit jeder Erhöhung um den Faktor 2,41. Ein SGB III Bezug 2008 lässt die Chance um 4,77 steigen, im Gegensatz zu keinen SGB III Bezug. Im Jahr 2007 nimmt die Chance eines Übergangs bei einem SGB III Bezug ab. Beim Alter ist dies ähnlich, die Chance nimmt mit jedem Lebensjahr um den Faktor 0,98 ab. Personen mit einem mittleren Abschluss haben eine 0,58 geringere Chance über zu gehen als Personen ohne Abschluss. Bei Personen mit einem hohen Abschluss sinkt der Faktor um 0,39. Im Gegensatz zu Chemnitz (Referenzkategorie) hat Meißen eine 0,69 geringere Chance einen Übergang zu haben. Dresden und Leipzig Stadt haben eine höhere Chance überzugehen. Dieser Faktor liegt in Dresden bei 1,14, in Leipzig Stadt liegt dieser bei 1,38. Die anderen Kreise haben eine geringe Chance als Chemnitz einen Übergang zu haben.

Die verschiedenen Modellgleichungen wirken sich folgendermaßen auf die Übergangswahrscheinlichkeiten in den Kreisen aus. Ausgegangen wird von dem anfänglichen Beispiel. Im 2. Schritt wird das Alter auf 60 Jahre erhöht.

Kreis	logit Analyse		ologit Analyse	
	Alter=25	Alter=60	Alter=25	Alter=60
Chemnitz	0,15	0,08	0,15	0,07
Erzgebirgskreis	0,12	0,06	0,12	0,06
Mittelsachsen	0,12	0,06	0,12	0,06
Vogtlandkreis	0,12	0,06	0,12	0,06
Zwickau	0,13	0,06	0,13	0,06
Dresden	0,17	0,09	0,17	0,08
Bautzen	0,13	0,06	0,12	0,06
Görlitz	0,14	0,07	0,14	0,07
Meißen	0,11	0,05	0,11	0,05
Sächsische Schweiz-Osterzgebirge	0,12	0,06	0,12	0,06
Leipzig Stadt	0,20	0,10	0,19	0,10
Leipzig	0,14	0,07	0,14	0,07
Nordsachsen	0,14	0,07	0,14	0,07
keine Angabe Kreis	0,11	0,05	0,10	0,05

Tabelle 4.17: Beispielrechnung für die beiden logistischen Modelle

In der deskriptiven Analyse wurde bereits erwähnt, dass häufig Personen aus Leipzig übergehen. Dies bestätigt sich. Es folgen Dresden und Chemnitz. Die niedrigste Übergangswahrscheinlichkeit hat Meißen mit nur 0,11. Bei den restlichen Kreisen sind die

Unterschiede als gering anzusehen. Es treten kaum Unterschiede auf. Ersichtlich werden die Unterschiede ab der 3. oder 4. Nachkommastelle. Bei der ologit Analyse wurde die Wahrscheinlichkeit berechnet, dass ein Übergang auftritt. Für einen Übergang in beiden Jahren sind die Wahrscheinlichkeiten sehr klein. Abschließend lässt sich feststellen, dass sowohl die binäre logistische Regression, wie auch die ordinale logistische Regression geeignet sind. Deshalb ist die binäre logistische Analyse als sinnvoller zu erachten.

5 Panelanalyse

Bei Paneldaten liegen zu bestimmten Zeitpunkten Daten vor, zum Beispiel die Zufriedenheit in jedem Jahr und für jede Person. Ist für jede Person in jedem Jahr ein Wert vorhanden, so nennt man das Panel „balanced“ andernfalls „unbalanced“³⁵. Das betrachtete Panel ist vom Typ „balanced“. [9] Der Vorteil derartiger Datenstrukturen liegt vor allem darin, dass sich Ereignisse im Verlauf der Zeit betrachtet werden können.

Das Panel wird eingeteilt in 2 Jahre. Einbezogen wurden die Ausprägungen:

Anzahl der Nebenbeschäftigungen im Jahr X
Anzahl der arbeitsmarktpolitischen Maßnahmen im Jahr x
Alter im Jahr X
Anzahl der Übergänge im Jahr X
höchster Abschluss im Jahr X
SGBIII Empfang im Jahr X ja/nein
Kreis im Jahr X

In einer Panelanalyse werden i Personen, an t Zeitpunkten beobachtet. Damit gibt es $N=i*t$ Beobachtungen. Ein Panel bei dem N größer ist als t nennt man Mikropanel, gilt der umgekehrte Fall so wird es als Makropanel bezeichnet. Nachteile der Panelanalyse sind.

- Die Beobachtungen einer Person in den verschiedenen Jahren sind voneinander abhängig.
- Panelmortalität³⁶ kann auftreten.
- Fehlende Daten, weil zum Beispiel keine Angaben gemacht wurden, führen ebenfalls zu verzerrten Ergebnissen.
- Es muss darauf geachtet werden, dass die Art der Befragung konstant bleibt. Dies ist gefährdet, wenn neue Fragen hinzukommen, sich die Bedeutung der Fragen ändert oder Fragen entfernt werden.

Panelmortalität kann in den betrachteten Datensatz nicht auftreten, da es keine Möglichkeit gibt, dass die Person nicht weiter geführt wird. Die Anzahl der fehlenden Werte ist ebenfalls als gering anzusehen. Die Problematik, dass sich Befragungen ändern, kann nur dann auftreten, wenn es Systemumstellungen gibt und dadurch Variablen hinzugefügt oder entfernt werden. Die Vorteile gegenüber anderen Regressionen.

³⁵ Es kommt häufig zu Problemen in der Auswertung

³⁶ Bezeichnet den Fall, dass Personen aus verschiedenen Gründen vorzeitig aus den Panel ausscheiden. Damit sind nicht für alle Jahre Beobachtungen vorhanden. Es kommt zu Verzerrungen im Panel.

- Es wird nicht nur die Varianz zwischen den einzelnen Personen beachtet, sondern auch die Varianz innerhalb der Person.
- Der Informationsgehalt eines Panels ist wesentlich größer.
- Es können zeitliche Entwicklungen verfolgt werden.

Entsprechend der Varianzvorgabe sind verschiedene Regressionen entwickelt worden, die entweder die Varianz zwischen oder innerhalb einer Person beachten. Bei der Betrachtung von Varianzen zwischen den Personen nennt sich dieses *between*. Varianzen innerhalb von Personen werden als *within* bezeichnet. Für die Verarbeitung müssen die Daten im Long Format vorliegen. Oft ist dies nicht der Fall und die Daten liegen im Wide Format vor. Der Datensatz wird deshalb transformiert. In Stata geschieht dies mit dem Befehl *reshape long vars, i(var) j(var)*. Wobei die vars die Variablen sind welche umgewandelt werden sollen, i gibt die Spalte an, welche die Ids der Personen enthält. In j wird die Variable geschrieben, in welche die Jahre gespeichert werden sollen. Beispiel für Daten im wide Format:

Personennummer	Alter 2007	Alter 2008	Geschlecht
1	27	28	männlich
2	53	54	männlich
3	31	32	weiblich

Tabelle 5.1: Daten im Wide Format

Nach der Umwandlung in das Long-Format entsteht folgende Tabelle:

Personennummer	Jahr	Alter	Geschlecht
1	2007	27	männlich
1	2008	28	männlich
2	2007	53	männlich
2	2008	54	männlich
3	2007	31	weiblich
3	2008	32	weiblich

Tabelle 5.2: Daten im Long-Format

Im schlimmsten Fall verdoppeln sich die Spalten in dem oberen Beispiel. Es wird davon ausgegangen, dass für jede Person zu jedem Zeitpunkt alle Werte vorhanden sind. Es gibt 2 Arten von Variablen. Variablen, welche über die Zeit gleich bleiben, zum Beispiel das Geschlecht. Variablen, welche sich im Laufe der Zeit ändern, sind zum Beispiel das Alter. Mittels des Befehls *tsset [Personenid] [jahr]* wird in Stata festgelegt, dass es sich bei den folgenden Berechnungen um Panelanalysen handelt. Es wird unterschieden in das Random Effects, Fixed Effects und das Hybrid Modell, eine Mischung aus Random und Fixed Effects Modell. Das Fixed Effects Modell (FE), betrachtet ausschließlich die Varianz innerhalb von Personen. Es können dabei keine zeitkonstanten Variablen beachtet werden. Im Random Effects Modell werden beide Varianzen berücksichtigt, außerdem können in diesem Modell zeitkonstante Variablen aufgenommen werden.

5.1 Fixed Effects Modell

Bei dem Fixed Effects Modell geht es darum, alle zeitkonstanten Personeneffekte zu eliminieren. Dies umfasst sowohl die beobachteten, wie auch die unbeobachteten Effekte. Gegeben sei folgende Gleichung aus Kapitel 4:

$$\ln \left(\frac{p(y=1)}{1-p(y=1)} \right) = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \dots + \beta_j x_{j,i} + u_i \quad (5.1)$$

Für die Berechnung des Fixed Effects Modells, kommt der personenspezifische Fehlerterm α_k dazu. Die unabhängigen Variablen werden außerdem von der Zeit abhängig gemacht. Auf den Fehlerterm u_{kt} wird im Folgenden verzichtet.

$$\ln \left(\frac{p(y=1)}{1-p(y=1)} \right) = \beta_0 + \alpha_i + \beta_1 x_{it1} + \beta_2 x_{it2} + \dots + \beta_j x_{itk} \quad (5.2)$$

Der Fehlerterm α_k wird aus der Gleichung eliminiert. Es wird auf die Conditional-Likelihood-Schätzung³⁷ zurückgegriffen. Dabei wird die bedingte Wahrscheinlichkeit maximiert, dass ein bestimmtes Ereignis eintritt. Das geschieht auf der Grundlage der Likelihoodfunktion, diese ist definiert als:

$$L = P(y_1 | x_{11}, \dots, x_{1k}) \dots P(y_n | x_{n1}, \dots, x_{nk}) \quad (5.3)$$

Bei den Berechnungen werden die Jahre 2007 und 2008 einbezogen, demnach gibt es 2 Wellen. Es wird die Summe der Ereignisse betrachtet. Folgende Kombinationen sind dabei möglich. 0 steht für keinen Übergang, 1 für einen Übergang.

$P(00 | \Sigma = 0) = 1$ damit tritt keine Veränderung von y ein

$P(01 | \Sigma = 1) < 1$ es tritt eine Veränderung von y ein

$P(10 | \Sigma = 1) < 1$ es tritt eine Veränderung von y ein

$P(11 | \Sigma = 2) = 1$ damit tritt keine Veränderung von y ein

Damit tragen Wahrscheinlichkeiten, in denen y keine Veränderung erfährt, nicht zu der Likelihoodfunktion bei, da dies apriori sichere Ereignisse sind. Für das Beispiel bleiben 2 Fälle übrig.

$$\frac{\frac{1}{1+e^{u_i+x'_{i1}\beta}} \cdot \frac{e^{u_i+x'_{i2}\beta}}{1+e^{u_i+x'_{i2}\beta}}}{\frac{1}{1+e^{u_i+x'_{i1}\beta}} \cdot \frac{e^{u_i+x'_{i2}\beta}}{1+e^{u_i+x'_{i2}\beta}} + \frac{e^{u_i+x'_{i1}\beta}}{1+e^{u_i+x'_{i1}\beta}} \cdot \frac{1}{1+e^{u_i+x'_{i2}\beta}}} = \frac{e^{u_i+x'_{i2}\beta}}{e^{u_i+x'_{i1}\beta} + e^{u_i+x'_{i2}\beta}} \quad (5.4)$$

Der Term e^{u_i} wird aus der Gleichung gekürzt. Damit entsteht die bedingte Wahrscheinlichkeit für das Ereignis 01.

$$\frac{e^{x'_{i2}\beta}}{e^{x'_{i1}\beta} + e^{x'_{i2}\beta}} \quad (5.5)$$

³⁷ auch bedingte Maximum-Likelihood-Schätzung

Durch Umformen erhält man folgende Wahrscheinlichkeit. Verglichen mit der Wahrscheinlichkeit einer logistischen Regression, wird die Ähnlichkeit deutlich.

$$\begin{aligned} Pr(01|\sum = 1) &= \frac{e^{x'_{i2}\beta}}{e^{x'_{i1}\beta} + e^{x'_{i2}\beta}} \cdot \frac{e^{-x'_{i1}\beta}}{e^{-x'_{i1}\beta}} = \frac{e^{(x'_{i2}-x'_{i1})\beta}}{1 + e^{(x'_{i2}-x'_{i1})\beta}} \\ Pr(y_i = 1) &= \frac{e^{x'_i\beta}}{1 + e^{x'_i\beta}} \end{aligned} \quad (5.6)$$

Die beschriebene Conditional-Likelihood-Schätzung wurde für $T=2$ durchgeführt. Ist T größer als 2, so werden die Kombinationen von Ereignissen angepasst. Die zeitkonstanten Terme fließen nicht mit in die Berechnung ein, dies geschieht durch den Fakt, dass Ereignisse ohne Veränderungen, die bedingte Wahrscheinlichkeit also 1 ist, nicht mit einfließen.

5.2 Random Effects Modell

Bei der Berechnung eines Random Effects Modell werden die zeitkonstanten Terme mit berücksichtigt. Hierbei muss die Voraussetzung gelten, dass die Fehlerterme u_i unkorreliert sind mit den unabhängigen Variablen x_{ikt} . In der Literatur wird meist zusätzlich angenommen, dass der Fehlerterm normalverteilt ist. Der Beitrag zu Likelihoodfunktion ergibt sich für jede Person i als.

$$L_i = P(y_{i1}, \dots, y_{iT}) = \int_{-\infty}^{+\infty} \left[\prod_t^T P(Y_{it} = y_{it} | x'_{it}\beta + u_i) \right] \cdot f(u_i) du_i \quad (5.7)$$

Der personenspezifische Fehlerterm wird durch die Dichtefunktion $f(u_i)$ dargestellt. Da angenommen wird, dass die Korrelation zwischen den unabhängigen Variablen und den Fehlerterm gleich 0 ist, kann die Likelihood als einfaches Produkt geschrieben werden.

Annahme 1:

Es wird angenommen, dass $u_i = \alpha_0$ gilt. Der Fehlerterm tritt damit bei jeder Person mit der Wahrscheinlichkeit 1 ein. Es wird von einer unabhängigen Variablen ausgegangen und 2 auftretenden Wellen.

$$L_1 = P(y_{11}, y_{12} | x_{111}, x_{112}) = [P(y_{11} | \alpha_0 + \beta_0 + \beta_1 x_{111}) \cdot P(y_{12} | \alpha_0 + \beta_0 + \beta_1 x_{112})] \cdot P(\alpha_0) \quad (5.8)$$

Der Fehlerterm wird in der Berechnung nicht auftauchen. Die β_0 sind nicht identifizierbar.

Annahme2:

Eine andere Annahme geht davon aus, dass verschiedene α_i mit unterschiedlichen

Wahrscheinlichkeiten auftraten. Dann gibt die Likelihoodfunktion die Wahrscheinlichkeit an das entweder α_1 , α_2 ... zutrifft.

$$\begin{aligned}
 l_1 = P(y_{11}, y_{12} | x_{111}, x_{112}) = & [P(y_{11} | \alpha_0 + \beta_0 + \beta_1 x_{111}) \cdot P(y_{12} | \alpha_0 + \beta_0 + \beta_1 x_{112})] \cdot P(\alpha_1) \\
 & + [P(y_{11} | \alpha_0 + \beta_0 + \beta_1 x_{111}) \cdot P(y_{12} | \alpha_0 + \beta_0 + \beta_1 x_{112})] \cdot P(\alpha_2) \\
 & + [P(y_{11} | \alpha_0 + \beta_0 + \beta_1 x_{111}) \cdot P(y_{12} | \alpha_0 + \beta_0 + \beta_1 x_{112})] \cdot P(\alpha_3) \\
 & \vdots \\
 & + [P(y_{11} | \alpha_0 + \beta_0 + \beta_1 x_{111}) \cdot P(y_{12} | \alpha_0 + \beta_0 + \beta_1 x_{112})] \cdot P(\alpha_n)
 \end{aligned}
 \tag{5.9}$$

Auch in diesen Fall sind die β_0 nicht identifizierbar.

Annahme3:

Hierbei wird davon ausgegangen das die α_i verschiedene Werte annehmen kann, welche zwischen $-\infty$ und $+\infty$ liegen können. Dabei wird der Fehlerterm als Dichtefunktion beschrieben.

$$\begin{aligned}
 l_1 = P(y_{11}, y_{12} | x_{111}, x_{112}) = & [P(y_{11} | \alpha' + \beta_0 + \beta_1 x_{111}) \cdot P(y_{12} | \alpha' + \beta_0 + \beta_1 x_{112})] \cdot f(\alpha') \\
 & + [P(y_{11} | \alpha'' + \beta_0 + \beta_1 x_{111}) \cdot P(y_{12} | \alpha'' + \beta_0 + \beta_1 x_{112})] \cdot f(\alpha'') \\
 & \vdots \\
 & + [P(y_{11} | \alpha^{(n)} + \beta_0 + \beta_1 x_{111}) \cdot P(y_{12} | \alpha^{(n)} + \beta_0 + \beta_1 x_{112})] \cdot f(\alpha^{(n)})
 \end{aligned}
 \tag{5.10}$$

Damit wird ersichtlich wie sich die Gleichung 5.7 daraus ergibt. Die Integration ist damit eine verallgemeinerte Addition.

5.3 Auswertung

Der SGB III Bezug in den beiden Jahren, wurde von Stata wegen der auftretenden Multikollinearität bei allen folgenden Berechnungen entfernt. Außerdem wird eine Jahresdummyvariable in die Berechnung mit einbezogen, diese erfasst die Trends der abhängigen Variablen.

5.3.1 Fixed Effects

Variable	Regressions- koeffizient B	Standardfehler	Waldstatistik	Signifikanz	EXP(B)
Jahr 2008	0,0064768	0,0095210	0,68	0,496	1,0064978
Nebenbeschäftigung	0,5549484	0,0089015	62,34	0,000	1,7418511
Maßnahme	0,4724692	0,0094129	50,19	0,000	1,6039498
Alter	-0,0250999				0,9752125

Tabelle 5.3: Modell für das Fixed Effect Modell

Mit jeder Beschäftigung steigt der Faktor, um 1,7 einen Übergang zu haben. Bezieht man die Jahresdummyvariable ein, so erkennt man das das Jahr nur einen geringen

Unterschied ausmacht. Demnach wäre es irrelevant ob mit einer logistischen Regression gerechnet wird oder eine Panelanalyse. Bei den Maßnahmen sieht es ähnlich aus, mit jeder Einheit um die sich die Maßnahmen erhöhen, erhöht sich die Chance eines Übergangs um 1,6. Mit jeder Maßnahme nimmt die Chance zu, einen Übergang zu haben. Je älter die Person um so geringer ist die Chance überzugehen.

5.3.2 Random Effects

Variable	Regressions- koeffizient B	Standardfehler	Waldstatistik	Signifikanz	EXP(B)
Jahr 2008	0,0547316	0,0093821	5,83	0,000	1,0562571
Nebenbeschäftigung	0,9447571	0,0048843	193,43	0,000	2,5721885
Maßnahme	0,7852161	0,0056929	137,93	0,000	2,1928808
Alter	-0,0170802	0,0003720	-45,92	0,000	0,9830648
mittlerer Abschluss	-0,2898881	0,0278891	-10,39	0,000	0,7483473
hoher Abschluss	-0,7152034	0,0300043	-23,84	0,000	0,4890926
keine Angabe Abschluss	-1,7806030	0,0638263	-27,90	0,000	0,1685365
Erzgebirgskreis	-0,1549064	0,0256346	-6,04	0,000	0,8564953
Mittelsachsen	-0,1621960	0,0267705	-6,06	0,000	0,8502745
Vogtlandkreis	-0,1358122	0,0283614	-4,79	0,000	0,8730066
Zwickau	-0,1094685	0,0265683	-4,12	0,000	0,8963104
Dresden	0,1216294	0,0251334	4,84	0,000	1,1293355
Bautzen	-0,1616541	0,0265667	-6,08	0,000	0,8507354
Görlitz	-0,0353016	0,0267110	-1,32	0,186	0,9653142
Meißen	-0,2829129	0,0288431	-9,81	0,000	0,7535854
Sächsische Schweiz Osterzgebirge	-0,1357953	0,0281101	-4,83	0,000	0,8730213
Leipzig Stadt	0,2537052	0,0248149	10,22	0,000	1,2887918
Leipzig	-0,0676239	0,0277834	-2,43	0,015	0,9346119
Nordsachsen	-0,0123969	0,0284981	-0,44	0,664	0,9876796
Keine Angabe Kreis	-0,4252284	0,0393695	-10,80	0,000	0,6536205
Konstante	-2,3734850	0,0366257	-64,80	0,000	0,0931555
/lnsig2u	-9,9827070	1,2377320	-12,41	-7,557	0,0000460
σ_{μ}	0,0067965	0,0042061	0,00	0,023	1,0068196
rho	0,0000140	0,0000174	e>4 e -1,25	0,000	1,0000140

Tabelle 5.4: Modell für das Random Effect Modell

Die Kreise Nordsachsen und Görlitz sind nicht signifikant. Mit jeder Nebenbeschäftigung nimmt das Risiko eines Überganges um 2,57 zu. Bei 3 Nebenbeschäftigungen wären dies 2,57³. Das Risiko für Maßnahmen nimmt nur geringfügig weniger ab. Personen mit einem mittleren Abschluss gehen 0,75-mal weniger über als Personen mit keinem Abschluss. Bei einem hohen Abschluss gehen nur halb so viele Personen über als bei Personen ohne einen Abschluss. Menschen aus den Kreis Meißen gehen mit einer 0,75 geringeren Chance über als aus der Referenzgruppe Chemnitz. Leipzig Stadt und Dresden haben eine höhere Chance über zu gehen als Personen aus Chemnitz. Die höchste Chance hat Leipzig Stadt mit 1,28. Alle anderen Kreise haben eine geringere Chance

über zu gehen. σ_u gibt dabei die Standardabweichung des personenbezogenen Fehlerterms an. ρ berechnet sich folgendermaßen.

$$\rho = \frac{(\sigma_u)^2}{(\sigma_u)^2 + (\sigma_e)^2} \quad (5.11)$$

5.3.3 Vergleich der Modelle

Als ungeeignet scheint das Modell der Fixed Effects zu sein. Es entsteht ein Informationsverlust, da die zeitkonstanten Variablen nicht geschätzt werden können, diese werden lediglich kontrolliert. Die besten Ergebnisse liefert das Random Effects Modell. Es wird verdeutlicht durch die geringeren Standardfehler. Eine Gegenüberstellung der Gütekriterien bestätigt die Vermutung.

	χ^2 -Wert	Loglikelihood	Freiheitsgrade
Fixed Effect	8402,65	−32119,59	3
Random Effect	68001,42	−177359,11	20

Tabelle 5.5: Vergleich der beiden Modelle

6 Zusammenfassung

Mithilfe dieser Diplomarbeit sollte herausgearbeitet werden, inwieweit es möglich ist, auf Grundlage der IEB eine Auswertung bezüglich der Wahrscheinlichkeiten eines Überganges zu bestimmen und diese auszuwerten. Dies sollte auf Grundlage der regionalen Merkmale stattfinden. Nach ersten Berechnungen zeigte es sich, dass es keinen signifikanten Zusammenhang zwischen Übergängen und regionalen Merkmalen gibt. Die Auswertung mittels logistischer Modelle brachte signifikante Ergebnisse, welche Verwendung finden können. Die Panelauswertung brachte aufgrund der wenigen Jahre, kein besseres Ergebnis. Eine Panelauswertung auf Hauptagenturebene brachte, ebenfalls nicht den gewünschten Zusammenhang. Für spätere Betrachtungen könnte eine Ereignisanalyse in Betracht gezogen werden. In diesem Falle kann die Wahrscheinlichkeit berechnet werden aus den Zustand SGB III Empfänger in den Zustand SGB II Empfänger zu wechseln. Aufgrund der Komplexität konnte diese hier nicht durchgeführt werden.

Anhang A: Ausprägungen der Variablen

A.1 Ausprägungen von der Variablen hoechsterab

Ausprägungen	
0	missing
1	Kein Hauptschulabschluss
2	Abschluss der Schule für Lern- oder Geistigbehinderte
3	Hauptschulabschluss
4	Qualifizierter Hauptschulabschluss
5	Mittlere Reife
6	Fachhochschulreife
7	Abitur/ Hochschulreife
11	Ohne Fachhochschulreife
12	Ohne Fachhochschul-/Hochschulabschluss
13	Fachhochschulabschluss (Diplom)
14	Hochschulabschluss (Diplom)

Tabelle A.1: Ausprägungen von hoechsterab

A.2 Ausprägungen von der Variablen hauptagentur

Variablenwerte	Ausprägungen
71	Annaberg
72	Bautzen
73	Chemnitz
74	Dresden
75	Leipzig
76	Oschatz
77	Pirna
78	Plauen
79	Riesa
92	Zwickau

Tabelle A.2: Ausprägungen von hauptagentur

A.3 Ausprägungen von der Variablen ieb_wo_gst_num

Wert	Ausprägungen	Wert	Ausprägungen	Wert	Ausprägungen
7101	Annaberg	7401	Dresden	7813	Oelsnitz
7105	Aue	7405	Radebeul	7817	Reichenbach
7109	Marienberg	7409	Radeberg	1901	Riesa
7113	Schwarzenberg	7501	Leipzig	7905	Großenhain
7117	Zschopau	7505	Borna	7909	Meißen
7201	Bautzen	7509	Delitzsch	9201	Zwickau
7205	Bischofswerda	7513	Eilenburg	9205	Glauchau
7209	Görlitz	7525	Geithain	9209	Hohenstein-Ernstthal
7213	Hoyerswerda	7601	Oschatz	9213	Stollberg
7217	Kamenz	7605	Döbeln	9217	Werdau
7221	Löbau	7609	Torgau	9221	Limbach-Oberfrohna
7225	Niesky	7613	Grimma	0	Fehlerhafte Werte
7229	Weißwasser	7617	Wurzen		
7233	Zittau	7701	Pirna		
7301	Chemnitz	7709	Dippoldiswalde		
7305	Flöha	7717	Freital		
7309	Hainichen	7721	Sebnitz		
7313	Rochlitz	7801	Plauen		
7317	Brand-Erbisdorf	7805	Auerbach		
7321	Freiberg	7809	Klingenthal		

Tabelle A.3: Ausprägungen von ieb_wo_gst_num

A.4 Ausprägungen von der Variablen ieb_wo_krs_num

14511	Chemnitz Stadt
14521	Erzgebirgskreis
14522	Mittelsachsen
14523	Vogtlandkreis
14524	Zwickau
14612	Dresden
14625	Bautzen
14626	Görlitz
14627	Meißen
14628	Sächsische Schweiz-Osterzgebirge
14713	Leipzig Stadt
14729	Leipzig
14730	Nordsachsen

Tabelle A.4: Ausprägungen von ieb_wo_krs_num

Anhang B: Übergänge in den Geschäftsstellen

Ausprägung	Anzahl der übergegangenen Personen			Anzahl der SGB III Empfänger			Anteil der übergegangenen SGB III Empfänger in %		
	2007	2008	2009	2007	2008	2009	2007	2008	2009
Annaberg	670	694	506	4447	4625	4239	15,07	15,01	11,94
Aue	557	582	495	3885	4214	4213	14,34	13,81	11,75
Marienberg	385	391	290	2706	2908	2772	14,23	13,45	10,46
Schwarzenberg	357	349	282	2049	2159	2140	17,42	16,16	13,18
Zschopau	259	264	167	2037	2201	1967	12,71	11,99	8,49
Bautzen	799	818	601	4980	5317	4926	16,04	15,38	12,20
Bischofswerda	269	250	207	1727	1954	1791	15,58	12,79	11,56
Görlitz	574	567	337	2504	2664	2483	22,92	21,28	13,57
Hoyerswerda	542	539	427	3103	3397	3316	17,47	15,87	12,88
Kamenz	424	445	347	3512	3795	3725	12,07	11,73	9,32
Löbau	579	533	391	3207	3323	3155	18,05	16,04	12,39
Niesky	222	243	166	1650	1763	1480	13,45	13,78	11,22
Weißwasser	318	277	183	2058	2135	1757	15,45	12,97	10,42
Zittau	615	613	436	3090	3308	2877	19,90	18,53	15,15
Chemnitz	1501	1484	1119	8198	9143	9036	18,31	16,23	12,38
Flöha	238	235	167	1842	1941	1759	12,92	12,11	9,49
Hainichen	410	420	361	2735	2921	2777	14,99	14,38	13,00
Rochlitz	406	380	314	2904	3120	2901	13,98	12,18	10,82
Freiberg	514	570	428	3701	4132	4174	13,89	13,79	10,25
Dresden	3097	3218	2256	16163	18208	17890	19,16	17,67	12,61
Radebeul	206	218	169	1729	1951	1997	11,91	11,17	8,46
Radeberg	166	169	137	1333	1467	1622	12,45	11,52	8,45
Leipzig	3392	3681	2761	15376	17121	16009	22,06	21,50	17,25
Borna	713	801	568	4316	4856	4546	16,52	16,50	12,49
Delitzsch	447	448	313	2610	2900	2791	17,13	15,45	11,21
Eilenburg	380	405	270	2226	2444	2309	17,07	16,57	11,69
Geithain	161	153	104	1071	1200	1044	15,03	12,75	9,96
Oschatz	344	391	252	2343	2544	2381	14,68	15,37	10,58
Döbeln	593	580	397	3217	3498	3202	18,43	16,58	12,40
Torgau	442	420	299	2652	2874	2529	16,67	14,61	11,82
Grimma	460	485	409	3306	3677	3594	13,91	13,19	11,38
Wurzen	371	377	302	2552	2811	2594	14,54	13,41	11,64
Pirna	728	689	454	4497	4789	4440	16,19	14,39	10,23
Dippoldiswalde	241	249	189	1861	2037	1950	12,95	12,22	9,69
Freital	468	539	398	3067	3349	3139	15,26	16,09	12,68
Sebnitz	278	321	207	1941	2220	2112	14,32	14,46	9,80
Plauen	694	689	564	4248	4645	4724	16,34	14,83	11,94
Auerbach	363	355	318	2698	2992	2954	13,45	11,86	10,77
Klingenthal	123	183	151	1317	1481	1425	9,34	12,36	10,60
Oelsnitz	166	184	150	1546	1665	1596	10,74	11,05	9,40
Reichenbach	304	321	222	1923	2062	1970	15,81	15,57	11,27
Riesa	467	425	304	2446	2609	2398	19,09	16,29	12,68
Großenhain	310	333	211	2124	2314	2218	14,60	14,39	9,51
Meißen	617	673	536	3986	4433	4330	15,48	15,18	12,38
Zwickau	746	804	665	3714	4040	4090	20,09	19,90	16,26
Glauchau	317	333	294	2018	2228	2262	15,71	14,95	13,00
Hohenstein-Ernstthal	304	336	224	2141	2284	2398	14,20	14,71	9,34
Stollberg	532	493	443	3813	4175	4212	13,95	11,81	10,52
Werdau	740	710	584	5279	5511	5584	14,02	12,88	10,46
Limbach-Oberfrohna	157	167	128	1008	1117	1150	15,58	14,95	11,13

Tabelle B.1: Übergänge und SGB III Personen nach Geschäftsstellen

Anhang C: Programmcode

Listing C.1: Quellcode Startdatei

```

* Kathleen Wiesner
* Version 10.0
discard
set more off

***** Hinweise
*****

* Bitte die Parameter genau festlegen!!!
* Außerdem bitte bei Bedarf die set memory Anweisung anpassen
* es werden verschiedene Datenbanken abgespeichert:
* "einzel" bezeichnet wieviele Personen es gibt (zählt Person nur einmal
  pro Jahr)
* "alle" bezeichnet alle Übergänge (Personen können mehrmals zählen pro
  Jahr)
* "sgbiii" bezeichnet alls sgbiii Empfänger eines Jahres

***** Parameter festlegen
*****

*----- Größe der Datenbank in
  Megabyte-----
* die nachfolgenden Daten werden benötigt um ungefähr den
  Speicherplatzbedarf fest zu legen, bei der Größe der datenbank
* die reale Größe abgeben!!! Der Bedarf wird automatisch runter gerechnet

global observat=0      //Ungefähr Anzahl der Spells (ganz Sachsen=0 oder
  Teilmenge für Testläufe)
local sicher=2         //muss nicht geändert werden, legt den
  Speicherplatzpuffer fest
                        //falls jedoch Fehlermeldungen ala "no more room" kommen, dann
                        diesen Faktor auf 3 oder 4 ändern

*----- Zeitraum festlegen
  -----
global beginn=td(01jan2007) //Beginn des Zeitraumes bitte immer am 1.
  Januar anfangen
global ende=td(31dec2009)  //Ende des Zeitraums bitte immer am 31.
  Dezember enden

*----- Nur Auswertung durchführen
  -----

local auswert=0        //Nur Auswertung anzeigen, von vorher erzeugter Datei
0=nein/1=ja

```

```
*----- Pfad zum Heimatverzeichnis
_____

global Pfad="D:\Datenaustausch\RFN\WiesnerK"

*----- Verzeichnis erzeugen
_____

local verzeichnis=0      // Soll das verzeichnis "Übergänge" im
    Heimatverzeichnis erstellt werden? 0=nein/1=ja
    // Dient nur der Übersichtlichkeit, weil viele Dateien erzeugt
    werden

*----- Pfad zur Datenbank
_____

local Pfad2="D:\Datenaustausch\RFN\WiesnerK\Rohdaten\A011729_ergebnis.dta"
local Pfad3="D:\Datenaustausch\RFN\WiesnerK\Rohdaten"

sysdir set OLDPLACE "$Pfad\Übergänge\ado"

*----- Bundesland
_____

global bundes=14          // zwischen 1 und 16 je nach Bundesland
    // 01=Schleswig Holstein
    // 02=Hamburg
    // 03=Niedersachsen
    // 04=Bremen
    // 05=Nordrhein_Westfalen
    // 06=Hessen
    // 07=Rheinland Pfalz
    // 08=Baden Württemberg
    // 09=Bayern
    // 10=Gemeinde Saarland
    // 11=Berlin
    // 12=Brandenburg
    // 13=Mecklenburg Vorpommern
    // 14=Sachsen
    // 15=Sachsen-Anhalt
    // 16=Thüringen
```

```

*****
***** Ab hier nichts mehr ändern
*****
*****
*****
*----- wenn nur Auswertung
*-----

if $observat==0 global obser1=55000000
else global obser1=$observat

local memdb=round((( $obser1*22*(60/22)+4*$obser1)/(1024*1024))*'sicher'+1,2)

if 'auswert'==1 {
    capture confirm file "$Pfad\Übergänge\alle$beginn$ende$obser1.dta"
    if !_rc local anw=1
    else local anw=0
    if 'anw'==1 {
        use "$Pfad\Übergänge\alle$beginn$ende$obser1.dta"
        do "$Pfad\Übergänge\ado\auswerten.do"
    }
    else di "Kann nicht auswerten, Datei nicht vorhanden."
}

else {
*----- Plausibilitäten
*-----

if 'verzeichnis'==1 mkdir "$Pfad\Übergänge\"

*heraus filtern welche dateien schon vorhanden sind, gegebenenfalls werden
diese erzeugt
capture confirm file "$Pfad\Übergänge\bitcode$beginn$ende$obser1.dta"
if !_rc local bit=1 //Bitzerlegung bereits vorhanden
else local bit=0 //Bitzerlegung nicht vorhanden

capture confirm file "$Pfad\Übergänge\leistung$beginn$ende$obser1.dta"
if !_rc local leist=1 //Ausgangsdatensatz vorhanden
else local leist=0 //Ausgangsdatensatz n. vorhanden

if 'leist'==1 & 'bit'==1 {
    local t=0
    local bitzer=0
}
else if 'leist'==1 & 'bit'==0 {
    local t=0
    local bitzer=1
}
else if 'leist'==0 & 'bit'==1 {
    local t=1
    local bitzer=0
}
}

```

```

else if 'leist'==0 & 'bit'==0 {
    local t=1
    local bitzer=1
}

capture confirm file 'Pfad2'
if _rc==601 {
    noisily di "Datenbank nicht gefunden (Pfad überprüfen)"
    exit
}

* Zeiträume überprüfen
if $beginn>$ende {
    noisily di "Beginn des Zeitraums liegt nach dem Ende des Zeitraums"
    exit
}
if year($beginn)<2007 {
    noisily di "Beginn darf nicht vor 2007 liegen (unsichere Daten)"
    exit
}
if month($beginn)!=1 | day($beginn)!=01 {
    noisily di "Beginn muss 1. Januar sein"
    exit
}
if month($ende)!=12 | day($ende)!=31 {
    noisily di "Ende muss 31. Dezember sein"
    exit
}
if $bundes<1 | $bundes>16 {
    noisily di "Bundesland falsch eingegeben"
    exit
}
local a=1
local b=100
local c='a'+int(('b'-'a'+1)*runiform())

*----- Vorspann
log using "$Pfad\Übergänge\log$beginn$ende$obser1'c'", text replace

noisily di "Startzeit"
noisily di "$S_TIME am $S_DATE"
noisily di "***** Berechnung gestartet
*****"
noisily di "Gewählte Optionen:"
noisily di "-----"
noisily di "Beginn des Beobachtungszeitraumes:" $beginn
noisily di "Ende des Beobachtungszeitraumes:" $ende

```

```

if 't'==1 noisily di "Ausgangsdatensatz nicht vorhanden für Zeitraum wird
erzeugt"
else noisily di "Ausgangsdatensatz vorhanden, kein erzeugen notwendig"
if 'bitzer'==1 noisily di "Bitzerlegung nicht vorhanden, wird vorgenommen"
else noisily di "Bitzerlegung bereits vorhanden, kein erzeugen notwendig"
noisily di "Gewähltes Verzeichnis:"
noisily di "Unterordner, Übergänge, wurde angelegt, alle Ergebnisse werden
dort gespeichert"

*-----Datensatz vorbereiten-----
if 't'==1 {
  if $obser1!=0 {
    clear
    set memory 'memdb'm
    use ieb_erw_stat_num ieb_geb_dat ieb_sex_id ieb_zug_gr ieb_abg_num
    ieb_sna_id ieb_sgb_ii_tr_art ieb_sbs_id ieb_staat_num ieb_bsb_id
    ieb_beruf_num ieb_berufstellg_num ieb_comb_quelle ieb_prs_id
    ieb_quelle ieb_wo_gst_num ieb_wo_aa_num ieb_wo_gem_num
    ieb_wo_krs_num ieb_ao_gst_num ieb_beg_epi ieb_end_epi if _n<$obser1
    using "'Pfad2'"
    drop if ieb_beg_epi<$beginn
    save "$Pfad\Übergänge\leistung$beginn$ende$obser1.dta"
  }
  else {
    clear
    set memory 'memdb'm
    use ieb_erw_stat_num ieb_geb_dat ieb_sex_id ieb_zug_gr ieb_abg_num
    ieb_sna_id ieb_sgb_ii_tr_art ieb_sbs_id ieb_staat_num ieb_bsb_id
    ieb_beruf_num ieb_berufstellg_num ieb_comb_quelle ieb_prs_id
    ieb_quelle ieb_wo_gst_num ieb_wo_aa_num ieb_wo_gem_num
    ieb_wo_krs_num ieb_ao_gst_num ieb_beg_epi ieb_end_epi using "'Pfad2'"
    drop if ieb_beg_epi<$beginn
    save "$Pfad\Übergänge\leistung$beginn$ende$obser1.dta"
  }
}

if 'bitzer'==0 {
  clear
  set memory 'memdb'm
  use "$Pfad\Übergänge\bitcode$beginn$ende$obser1.dta"

  // weiteren Speicherplatzbedarf berechnen, zur Sicherheit werden 20% für
  // neue Variablen usw. dazu gegeben
  sum
  local gesamt=r(N)
  local
  mem=round((( 'gesamt'*28*(62/28)+4*'gesamt')/(1024*1024))*'sicher'+1,2)
  clear
  set memory 'mem'm
  use "$Pfad\Übergänge\bitcode$beginn$ende$obser1.dta"

  local anz1=97397 //Anzahl des Referenz Durchlaufes

```

```

    local t2=36          //Zeit die vorraussichtlich benötigt wird ohne
                          Bitzerlegung in Sekunden
    sum
    local gesamt=r(N)
    noi di r(N)
    local tgesamt=(r(N)*'t2'/'anz1')/60
    noisily di "Vorraussichtliche Zeit bis zum fertig stellen in Minuten:
               "'tgesamt'"
}
else if 'bitzer'==1 {
    clear
    set memory 'memdb'm
    use "$Pfad\Übergänge\leistung$beginn$ende$obser1.dta"
    sum
    local gesamt=r(N)
    local
        mem1=round((( 'gesamt'*28*(62/28)+4*'gesamt')/(1024*1024))*'sicher'+1,2)
    clear
    set memory 'mem1'm
    use "$Pfad\Übergänge\leistung$beginn$ende$obser1.dta"

    local anz2=97515      //Anzahl des Referenz Durchlaufes'
    local t1=107          //Zeit die vorraussichtlich benötigt wird mit
                          Bitzerlegung in Sekunden
    sum
    noisily di r(N)
    local gesamt=r(N)
    noi di r(N)
    local tgesamt=(r(N)*'t1'/'anz2')/60
    noisily di "Vorraussichtliche Zeit bis zum fertig stellen in Minuten:
               "'tgesamt'"
}

*-----

global jahrbeginn=year($beginn)
global jahrende=year($ende)
global anzjahr=$jahrende-$jahrbeginn+1
noisily di "Anzahl der jahre=$anzjahr

if 'bitzer'==1 {
    * Intcount anlegen, zählt die Spells einer Person
    by ieb_prs_id, sort: generate intcount=_N

    * Spalte "Beginn Episode" verschieben damit leichtere Auswertung
    move ieb_beg_epi ieb_quelle
    move ieb_end_epi ieb_quelle

    * Sortieren nach der Kontoid und dann nach dem Beginn der Episode
    sort ieb_prs_id ieb_beg_epi

```



```

* Vorbereitungen
*-----
generate status=0
move status ieb_beg_epi
generate Zeitraumepi=ieb_end_epi-ieb_beg_epi+1
*-----

* ieb_comb_quelle in Bitcode zerlegen
* einfach Zuordnung wo welche Quelle einfließt
noisily di "Fange mit Bitcodezerlegung an..."
quietly bitzerlegung
noisily di "Bitcodezerlegung abgeschlossen"

compress
capture confirm file "$Pfad\Übergänge\bitcode$beginn$ende$obser1.dta"
if _rc==601 save "$Pfad\Übergänge\bitcode$beginn$ende$obser1.dta"
else save "$Pfad\Übergänge\bitcode$beginn$ende$obser1.dta", replace

sum
local gesamt=r(N)
local
    mem=round((( 'gesamt'*28*(62/28)+4*'gesamt')/(1024*1024))*'sicher'+1,2)
clear
set memory 'mem'm
use "$Pfad\Übergänge\bitcode$beginn$ende$obser1.dta"
}

datensatzordnen

compress

capture confirm file "$Pfad\Übergänge\alle$beginn$ende$obser1.dta"
if _rc==601 save "$Pfad\Übergänge\alle$beginn$ende$obser1.dta"
else save "$Pfad\Übergänge\alle$beginn$ende$obser1.dta", replace
noisily di "Datenbank erstellt"

sum
local gesamt=r(N)
local
    mem3=round((( 'gesamt'*28*(62/28)+4*'gesamt')/(1024*1024))*'sicher'+1,1)
clear
set memory 'mem3'm
use "$Pfad\Übergänge\alle$beginn$ende$obser1.dta"

log close

run "$Pfad\Übergänge\ado\auswerten.do"

```

```

noisily di "Alles fertig , logs befinden sich im Verzeichnis , können nun in
Excel eingelesen werden"

noisily di "Endzeit"
noisily di "$S_TIME"
noisily log close

}

```

Listing C.2: Quellcode zum ordnen des Datensatzes

```

program datensatzordnen

*****
*Zuordnung der Leistungsart*
*****

generate uebersicht=0

replace status=1 if (Leh==1 & Lhg!=1 & Lhgx!=1)
replace status=2 if (Leh!=1 & Lhg==1 & Lhgx!=1) | (Leh!=1 & Lhg!=1 &
    Lhgx==1) | (Leh!=1 & Lhg==1 & Lhgx==1)
replace status=3 if (Leh==1 & Lhg!=1 & Lhgx==1) | (Leh==1 & Lhg==1 &
    Lhgx!=1)
replace uebersicht=3 if status==1
replace uebersicht=2 if status==2 | status==3
label define status 1 "SGBIII Bezug" 2 "SGBII Bezug" 3 "SGBIII & SGBII
    Bezug" 0 "Keine Leistung"
label values status status

drop intcount
by ieb_prs_id ieb_beg_epi, sort: gen intcount=_N

*****
*Zuordnung ob nebenbei eine Beschäftigung oder Massnahme aufgeführt wird*
*****

gen statepi=0
local z=1
local zaehler=0
local co=0
local spells=0
while 'spells'<_N {
    local z='zaehler'+`z'
    local zaehler=intcount[`z']
    local y='zaehler'+`z'-1
    if 'zaehler'>1 {
        forvalues co=`z'/'y' {
            replace statepi=1 if ieb_erw_stat_num[`co']>=10101 &
                ieb_erw_stat_num[`co']<=27118 in `z'/'y'
        }
    }
}

```

```

        replace statepi=2 if ieb_erw_stat_num['co']>100 &
            ieb_erw_stat_num['co']<210 in 'z'/'y'
    }
}
if 'zaehler'==1 {
    replace statepi=ieb_erw_stat_num['z'] in 'z'
}
local spells='spells'+ 'zaehler'
}
replace statepi=1 if statepi>=10101 & statepi<=27118
replace statepi=2 if statepi>100 & statepi<210
replace statepi=0 if statepi>2
replace statepi=3 if status==0

label define statussgbiii 0 "ohne Alles" 1 "Massnahme" 2 "zusätzl. Arbeit"
label values statepi statussgbiii

*****
*fehlende Werte bei Abschluss und Behinderung füllen, dabei werden die *
*gesamten Spells einer Person durchsucht, ist ein Wert >0 so wird dieser*
*in allen Zeilen der Person ersetzt                                     *
*****

mvencode ieb_bsb_id, mv(0)
mvencode ieb_sbs_id, mv(0)
replace ieb_bsb_id=0 if ieb_bsb_id==9997
replace ieb_sbs_id=0 if ieb_sbs_id==9997

local z=1
local zaehler=0
local co=0
local spells=0
while 'spells'<_N {
    local z='zaehler'+ 'z'
    local zaehler=intcount['z']
    local y='zaehler'+ 'z'-1
    if 'zaehler'>1 {
        forvalues co='z'/'y' {
            replace ieb_bsb_id=ieb_bsb_id['co'] if ieb_bsb_id['co']>0 in 'z'/'y'
            replace ieb_sbs_id=ieb_sbs_id['co'] if ieb_sbs_id['co']>0 in 'z'/'y'
        }
    }
}
local spells='spells'+ 'zaehler'
}

drop intcount
by ieb_prs_id ieb_beg_epi, sort: gen intcount=_n
drop if intcount>1

*****
*Stringzerlegung damit man heraus finden kann wer aus Sachsen kommt *

```

```
*****
```

```
generate sachsen=0
tostring ieb_wo_gem_num, replace    //in String Variable umwandeln
local j=_N
forvalues co=1/'j' {                //String zerlegen
    local wort=ieb_wo_gem_num['co']
    local k=substr("`wort'",1,2)
    replace sachsen='k' in 'co'
    local co='co'+1
}
destring sachsen, replace
```

```
drop intcount
by ieb_prs_id , sort: gen intcount=_N
```

```
*****
*alle die nicht aus Sachsen kommen werden gelöscht*
*****
```

```
generate leben=0
```

```
local z=1
local zaehler=0
local count=0
while 'count'<_N{
    local z='zaehler'+`z'
    local zaehler=intcount['z']
    local y='zaehler'+`z'-1
    if 'zaehler'>1 {
        count if sachsen==14 in 'z'/'y'
        local a=r(N)
        replace leben=1 if 'a'=='zaehler' in 'z'/'y' //keine Markierung
        außerhalb
        replace leben=2 if 'a'==0 in 'z'/'y' //keine Markierung in Sachsen
    }
    local count='count'+`zaehler'
}
drop if leben==2
```

```
*****
*Der nächste Algorithmus addiert nacheinander folgende Zeiträume auf*
*Dies geschieht immer wenn der Zeitraum 1 Tag beträgt und der Status*
*gleich geblieben ist*
*****
```

```
gen Zeitraum=0
```

```
local j=_N
forvalues co=2/'j' {
```

```

    local l='co'-1
    local spanne=ieb_beg_epi['co']-ieb_end_epi['l']
    replace Zeitraum='spanne' in 'co'
}
drop intcount
by ieb_prs_id, sort: gen intcount=_N

gen aufadd=0

local z=1
local zaehler=0
local co=0
local spells=0
replace aufadd=0
while 'spells'<_N {
    local z='zaehler'+ 'z'
    local zaehler=intcount['z']
    local y='zaehler'+ 'z'-1

    local stat=0
    local zeit=0
    local abm=0

    if 'zaehler'>1 {
        forvalues co='z'/'y' {
            local l='co'+1
            if l<='y' & aufadd['co']==0{
                local k=1
                local stat=status['co']
                local zeit=Zeitraum['co']
                local abm=statepi['co']
                local stelle='co'
                while (status['l']== 'stat' & statepi['l']== 'abm' &
                    Zeitraum['l']==1) {
                    local k='k'+1
                    local co='co'+1
                    local l='l'+1
                }
                local stelle2='stelle'+ 'k'-1
                replace aufadd='k' in 'stelle'/'stelle2'
            }
        }
    }
    local spells='spells'+ 'zaehler'
}

replace aufadd=1 in 1

generate aussort=0

replace aussort=0
local z=1

```

```

local zaehler=0
local spells=0
while 'spells'<_N {
  local z='zaehler'+ 'z'
  local zaehler=aufadd[ 'z' ]
  local y='zaehler'+ 'z'-1
  if 'zaehler'>0 {
    replace aussort=1 in 'z'
    replace ieb_end_epi=ieb_end_epi[ 'y' ] in 'z'
  }
  else if 'zaehler'==0 local zaehler=1
  local spells='spells'+ 'zaehler'
}

replace aussort=1 if aufadd==0

drop if aussort==0

drop intcount
by ieb_prs_id, sort: gen intcount=_N

*****
*Dieser Algorithmus markiert die Übergänge einer Person, dabei steht 1 *
*für das Ende des SGBIII Leistungsbezuges und eine 2 für den Anfang des*
*SGBII Leistungsbezuges                                     *
*****

generate uebergaenge=0

replace uebergaenge=0
local z=1
local zaehler=0
local spells=0
local m=0          //speichert Wert wenn SGBIII Anspruch
local k=0
local co=0
local l=0          //speichert Stelle wo SGBIII Anspruch auftritt
while 'spells'<_N {
  local z='zaehler'+ 'z'
  local zaehler=intcount[ 'z' ]
  local y='zaehler'+ 'z'-1
  local m=0
  local k=0
  local j=0
  local p=0
  forvalues co='z'/'y' {
    if uebersicht[ 'co' ]==3 {
      local m=3
      local k='co'
      local j='co'
    }
  }
}

```

```

else if uebersicht['co']==2 & 'm'==3 {
  if 'k'=='j' {
    local p='j'+1
    replace uebergaenge=1 in 'j'
    replace uebergaenge=2 in 'p'
    local k=0
    local m=0
    local j=0
    local p=0
  }
  else {
    replace uebergaenge=1 in 'j'
    replace uebergaenge=2 in 'co'
    local k=0
    local m=0
    local j=0
  }
}
else if uebersicht['co']==0 & 'm'==3 {
  local k='k'+1
}

local co='co'+1
}
local spells='spells'+ 'zaehler'
}

*****
*Leute entfernen die nie SGBIII bekommen haben, könne auch nicht übergehen*
*****

replace aussort=0
local z=1
local zaehler=0
local spells=0
while 'spells'<_N{
  local z='zaehler'+ 'z'
  local zaehler=intcount['z']
  local y='zaehler'+ 'z'-1
  local i=0
  if 'zaehler'>1 {
    forvalues co='z'/'y' {
      if status['co']!=1 local i='i' +1
    }
    if 'i'=='zaehler' replace aussort=1 in 'z'/'y'
  }
  local spells='spells'+ 'zaehler'
}

drop if aussort==1

```

```

*****
*Übergänge außerhalb von Sachsen löschen*
*****

drop intcount
by ieb_prs_id, sort: gen intcount=_N
local z=1
local zaehler=0
local spells=0
while 'spells'<_N {
    local z='zaehler'+`z'
    local zaehler=intcount[ `z' ]
    local y='zaehler'+`z'-1
    local i=0
    forvalues co=`z'/'y' {
        if (uebergaenge[ `co' ]==1 & sachsen[ `co' ]!=14) replace aussort=1 in
            `z'/'y'
    }
    local spells='spells'+`zaehler'
}
drop if aussort==1

drop if ieb_end_epi>td(31dec2009)

generate zuordjahrbeg=0
generate zuordjahrend=0
generate zuordmonat=0

replace zuordjahrbeg=year(ieb_beg_epi)
replace zuordjahrend=year(ieb_end_epi)
replace zuordmonat=month(ieb_end_epi) if uebergaenge==1

global jahrbeginn=year($beginn)
global jahrende=year($ende)

forvalues jahr=$jahrbeginn/$jahrende {
    generate jahr `jahr'=0
}

local spanne=$jahrende-$jahrbeginn

if 'spanne'==0 {
    replace jahr$jahrbeginn=1 if zuordjahrbeg!=.
}
else {
    local count=1
    local gesamt=_N
    forvalues count=1/'gesamt' {

```



```

    if zuordjahrbeg['count']<10000 {
        local jahr1=zuordjahrbeg['count']
        local jahr2=zuordjahrend['count']
        forvalues z='jahr1'/'jahr2' {
            replace jahr'z'=1 in 'count'
        }
    }
    local count='count'+1
}
}

*****
*****
**Die Daten für die Auswertung vorbereiten**
*****
*****

*****
*Herausfiltern des höchsten Abschlusses*
*****

generate hoechsterab=0
recast int hoechsterab
local z=1
local zaehler=0
local co=0
local l=0
local spells=0
while 'spells'<_N {
    local z='zaehler'+ 'z'
    local zaehler=intcount['z']
    local y='zaehler'+ 'z'-1
    forvalues co='z'/'y' {
        if ieb_bsb_id['co']==0 {
            local co='co'+1
        }
        else if ieb_bsb_id['co']!=0 & 'l'==0{
            local l=ieb_bsb_id['co']
            local co='co'+1
        }
        else if ieb_bsb_id['co']!=0 & 'l'!=0{
            if ieb_bsb_id['co']>'l' local l=ieb_bsb_id['co']
        }
    }
    replace hoechsterab='l' in 'z'/'y'
    local l=0
    local spells='spells'+ 'zaehler'
}

*****

```

```
*Herausfiltern der Schwerbehinderung      *
*****
```

```
generate schwerbe=0
recast int schwerbe
local z=1
local zaehler=0
local co=0
local spells=0
while 'spells'<_N {
    local z='zaehler'+`z'
    local zaehler=intcount[ `z' ]
    local y='zaehler'+`z'-1
    forvalues co=`z'/'y' {
        if ieb_sbs_id[ `co' ]==0 {
            local co=`co'+1
        }
        else if ieb_sbs_id[ `co' ]!=0 {
            replace schwerbe=ieb_sbs_id[ `co' ] in `z'/'y'
            local co=`y'
        }
    }
    local spells='spells'+`zaehler'
}

```

```
label values hoechsterab BSB_DIL
```

```
*****
*Bestimmen der Pause zwischen SGBIII und SGBII*
*****
```

```
generate laengepause=0
local s=0
local co =0

while 'co'<=_N {

    if uebergaenge[ `co' ]==1 {
        local sgb3=ieb_end_epi[ `co' ]
        local s=`co'
    }
    else if uebergaenge[ `co' ]==2 & ieb_prs_id[ `s' ]==ieb_prs_id[ `co' ] {
        local sgb2=ieb_beg_epi[ `co' ]
        local laenge='sgb2'-'sgb3'
        replace laengepause='laenge' in `s'
        local s=0
    }

    local co=`co'+1
}

```

```
label variable hoechsterab "Abschluss"
label variable schwerbe "Behinderung"
```

```
*****
*Bestimmen verschiedener Werte im wide-Format, später erfolgt Änderung ins
  long-Format*
*****
```

```
gen arbeit2007=0          //Hatte die Person im Jahr X nebenbei
                           gearbeitet, wenn ja wieviele
gen arbeit2008=0
gen arbeit2009=0
gen abm2007=0             //Hatte die Person im Jahr X eine arbeitsmarktpol.
                           Massnahme, wenn ja wieviele
gen abm2008=0
gen abm2009=0
gen ueb=0                 //Anzahl der Übergänge der Person
generate direkt=0         //Markierung ob unter 7 Tage oder über 7 Tage
                           Pause zwischen Leistungen
replace laengepause=. if laengepause==0
gen alter2007=0           //Alter im Jahr X
gen alter2008=0
gen alter2009=0
gen sgb2007=0             //Hatte die Person im Jahr X einen SGBIII Bezug
gen sgb2008=0
gen sgb2009=0
gen ueb2007=0            //Hatte die Person im Jahr X einen Übergang
gen ueb2008=0
gen ueb2009=0
```

```
local z=1
local zaehler=0
local spells=0
while 'spells'<_N {
    local z='zaehler'+`z'
    local zaehler=intcount[ `z' ]
    local y='zaehler'+`z'-1
    local i=0
    local j=0
    local k=0
    local l=0
    local m=0
    local n=0
    local o=0
    forvalues co=`z'/'y' {
        if statepi[ `co' ]==1 & jahr2007[ `co' ]==1 local i=`i'+1
        if statepi[ `co' ]==1 & jahr2008[ `co' ]==1 local j=`j'+1
        if statepi[ `co' ]==1 & jahr2009[ `co' ]==1 local k=`k'+1
        if statepi[ `co' ]==2 & jahr2007[ `co' ]==1 local l=`l'+1
        if statepi[ `co' ]==2 & jahr2008[ `co' ]==1 local m=`m'+1
```

```

if statepi['co']==2 & jahr2009['co']==1 local n='n'+1
replace direkt=1 in 'z' if laengepause['co']<=7
replace sgb2007=1 in 'z' if status['co']==1 & jahr2007['co']==1
replace sgb2008=1 in 'z' if status['co']==1 & jahr2008['co']==1
replace sgb2009=1 in 'z' if status['co']==1 & jahr2009['co']==1
replace ueb2007=1 in 'z' if uebergaenge['co']==1 & jahr2007['co']==1
replace ueb2008=1 in 'z' if uebergaenge['co']==1 & jahr2008['co']==1
replace ueb2009=1 in 'z' if uebergaenge['co']==1 & jahr2009['co']==1
if uebergaenge['co']==1 local o='o'+1
}
replace arbeit2007='l' in 'z'
replace arbeit2008='m' in 'z'
replace arbeit2009='n' in 'z'
replace abm2007='i' in 'z'
replace abm2008='j' in 'z'
replace abm2009='k' in 'z'
replace ueb='o' in 'z'
local spells='spells'+ 'zaehler'
}

*****
*Zusammenfassen zu Kategorien, da viele Kategorien unterbesetzt sind*
*Außerdem Vergabe von Label *
*****

replace hoechsterab=2 if hoechsterab==2 | hoechsterab==3 | hoechsterab==4|
    hoechsterab==5
replace hoechsterab=3 if hoechsterab==6 | hoechsterab==7 | hoechsterab==12|
    hoechsterab==13 | hoechsterab==14

replace ieb_staat_num=1 if ieb_staat_num>0
replace ieb_staat_num=9997 if ieb_staat_num<0

replace ieb_sex_id=0 if ieb_sex_id==2

replace schwerbe=1 if schwerbe==2 | schwerbe==3
replace schwerbe=2 if schwerbe==4

label define schwerbe 2 "nichts behindert" 1 "Behindert"
label values schwerbe schwerbe
label define ieb_staat_num 1 "ausländisch" 0 "deutsch"
label value ieb_staat_num ieb_staat_num
label define hoechsterab 1 "kein Abschluß" 2 "Mittlerer Abschluß" 3 "Hoher
    Abschluß"
label define geschlecht 0 "weiblich" 1 "männlich"
label value ieb_sex_id geschlecht
numlabel _all ,add

```

```

label define kreis 14511 "Chemnitz Stadt" 14521 "Erzgebirgskreis" 14522
    "Mittelsachsen" 14523 "Vogtlandkreis" 14524 "Zwickau" 14612 "Dresden"
    14625 "Bautzen" 14626 "Görlitz" 14627 "Meißen" 14628 "Sächsische
    Schweiz–Osterzgebirge" 14713 "Leipzig Stadt" 14729 "Leipzig" 14730
    "Nordsachsen"
label value ieb_wo_krs_num kreis
numlabel _all ,add

*****
*** ieb_wo_krs_num nach ganz alter ieb_wo_krs_numstruktur (Sachsen)
    umschlüsseln ***
*****

replace ieb_wo_krs_num = 14178 if ieb_wo_krs_num == 14045 // Plauen ->
    Vogtlandieb_wo_krs_num
replace ieb_wo_krs_num = 14365 if ieb_wo_krs_num == 14065 // Leipzig , Stadt
replace ieb_wo_krs_num = 14171 if ieb_wo_krs_num == 14071 // Annaberg
replace ieb_wo_krs_num = 14173 if ieb_wo_krs_num == 14073 // Chemnitzer Land
replace ieb_wo_krs_num = 14374 if ieb_wo_krs_num == 14074 // Delitzsch
replace ieb_wo_krs_num = 14379 if ieb_wo_krs_num == 14079 // Leipziger Land
replace ieb_wo_krs_num = 14181 if ieb_wo_krs_num == 14081 // mittl.
    Erzgebirgsieb_wo_krs_num
replace ieb_wo_krs_num = 14182 if ieb_wo_krs_num == 14082 // Mittweida
replace ieb_wo_krs_num = 14284 if ieb_wo_krs_num == 14084 // Niederschl.
    Oberlausitz
replace ieb_wo_krs_num = 14285 if ieb_wo_krs_num == 14085 //
    Riesa–Grossenhain
replace ieb_wo_krs_num = 14287 if ieb_wo_krs_num == 14087 // Saechsische
    Schweiz
replace ieb_wo_krs_num = 14290 if ieb_wo_krs_num == 14090 //
    Weißeritzieb_wo_krs_num

*****
*** Umschlüsselung der ieb_wo_krs_num zum zum 01.08.2008 (Sachsen) ***
*****

replace ieb_wo_krs_num = 14511 if ieb_wo_krs_num == 14161
replace ieb_wo_krs_num = 14521 if ieb_wo_krs_num == 14188 | ieb_wo_krs_num
    == 14191 | ieb_wo_krs_num == 14181 | ieb_wo_krs_num == 14171
replace ieb_wo_krs_num = 14522 if ieb_wo_krs_num == 14182 | ieb_wo_krs_num
    == 14177 | ieb_wo_krs_num == 14375
replace ieb_wo_krs_num = 14523 if ieb_wo_krs_num == 14166 | ieb_wo_krs_num
    == 14178
replace ieb_wo_krs_num = 14524 if ieb_wo_krs_num == 14167 | ieb_wo_krs_num
    == 14173 | ieb_wo_krs_num == 14193
replace ieb_wo_krs_num = 14612 if ieb_wo_krs_num == 14262
replace ieb_wo_krs_num = 14625 if ieb_wo_krs_num == 14272 | ieb_wo_krs_num
    == 14264 | ieb_wo_krs_num == 14292
replace ieb_wo_krs_num = 14626 if ieb_wo_krs_num == 14286 | ieb_wo_krs_num
    == 14284 | ieb_wo_krs_num == 14263
replace ieb_wo_krs_num = 14627 if ieb_wo_krs_num == 14280 | ieb_wo_krs_num
    == 14285

```

```

replace ieb_wo_krs_num = 14628 if ieb_wo_krs_num == 14290 | ieb_wo_krs_num
    == 14287
replace ieb_wo_krs_num = 14713 if ieb_wo_krs_num == 14365
replace ieb_wo_krs_num = 14729 if ieb_wo_krs_num == 14379 | ieb_wo_krs_num
    == 14383
replace ieb_wo_krs_num = 14730 if ieb_wo_krs_num == 14374 | ieb_wo_krs_num
    == 14389

replace ieb_wo_krs_num = 0 if ieb_wo_krs_num<14511 | ieb_wo_krs_num>14730

gen zuordjahr=0
replace zuordjahr=year(ieb_end_epi) if uebergaenge==1

tostring ieb_wo_gst_num, replace
generate hauptagentur=.
recast int hauptagentur
local j=_N
forvalues co=1/'j' {
    local wort=ieb_wo_gst_num['co']
    local k=substr("`wort'",1,2)
    replace hauptagentur='k' in 'co'
    local co='co'+1
}
destring ieb_wo_gst_num, replace

end
exit

```

Listing C.3: Quellcode für die Deskriptive Analyse

```

log using "$Pfad\Übergänge\logalle$beginn$ende", text replace

qui replace laengepause=8 if laengepause>=8 & laengepause<=28
qui replace laengepause=9 if laengepause>28
qui replace laengepause=. if laengepause==0

qui label define pause 1 "1_Tag" 2 "2_Tage" 3 "3_Tage" 4 "4_Tage" 5
    "5_Tage" 6 "6_Tage" 7 "1_Woche" 8 "1-4_Wochen" 9 ">4_Wochen"
qui label value laengepause pause

qui replace ieb_wo_gst_num=07321 if ieb_wo_gst_num==07317

qui label define schwer 9997 "Missing" 1 "anerkannt" 2 "gleichgestellt" 3
    "Gleichstellung_möglich" 4 "nicht_schwerbehindert"
qui label value schwerbe schwer

qui label drop BSB_DIL

```

```

qui label define BSB_DIL 1 "kein_hauptschulabschluss" 2
    "abschluss_der_schule_für_lern-oder_geistigbehinderte" 5
    "mittlere_Reife" 4 "qualifizierter_hauptschulabschluss" 11
    "ohne_fachhochschulreife" 12 "ohne_fachhochschul-/hochschulabschluss" 13
    "fachhochschulabschluss_(diplom)" 14 "hochschulabschluss_(diplom)" 9997
    "Fehler"
qui label value hoechsterab BSB_DIL

replace hauptagentur=999 if hauptagentur<71 | (hauptagentur>79 &
    hauptagentur<92) | hauptagentur>92

label define hauptagentur 999 "anderes Bundesland" -9 "Missing" 70
    "Altenberg" 71 "Annaberg" 72 "Bautzen" 73 "Chemnitz" 74 "Dresden" 75
    "Leipzig" 76 "Oschatz" 77 "Pirna" 78 "Plauen" 79 "Riesa" 92 "Zwickau"
label values hauptagentur hauptagentur

qui numlabel _all ,remove

replace ieb_wo_gst_num=99999 if ieb_wo_gst_num<0
replace ieb_wo_gst_num=999 if ieb_wo_gst_num<07101 |
    (ieb_wo_gst_num>09221 & ieb_wo_gst_num<99999)

replace zuordjahr=. if zuordjahr==0
replace zuordjahr=zuordjahrend if uebergaenge==1

di year($beginn)
di year($ende)

noi ta hauptagentur zuordjahr
noi ta ieb_wo_krs_num zuordjahr
noi ta ieb_wo_gst_num zuordjahr
noi ta schwerbe zuordjahr
noi ta ieb_staat_num zuordjahr
noi ta hoechsterab zuordjahr
noi ta ieb_sex_id zuordjahr
noi ta alter2007 zuordjahr
noi ta zuordmonat zuordjahr
noi ta ueb zuordjahr
noi ta laengepause zuordjahr

log close

log using "$Pfad\Übergänge\logeinzeln$beginn$ende", text replace

noi di year($beginn)
noi di year($ende)

forvalues jahr=2007/2009 {
    replace ueb'jahr'=. if ueb'jahr'==0
    noi ta hauptagentur ueb'jahr'
    noi ta ieb_wo_krs_num ueb'jahr'

```

```
noi ta ieb_wo_gst_num ueb'jahr'
noi ta schwerbe ueb'jahr'
noi ta ieb_staat_num ueb'jahr'
noi ta hoechsterab ueb'jahr'
noi ta ieb_sex_id ueb'jahr'
noi ta alter2007 ueb'jahr'
noi ta zuordmonat ueb'jahr'
noi ta ueb ueb'jahr'
noi ta laengepause ueb'jahr'
}
```

```
log close
```

```
log using "$Pfad\Übergänge\logsgbiii$beginn$ende", text replace
```

```
noi di year($beginn)
noi di year($ende)
```

```
forvalues jahr=2007/2009 {
  replace sgb'jahr'=. if sgb'jahr'==0
  noi ta hauptagentur sgb'jahr'
  noi ta ieb_wo_krs_num sgb'jahr'
  noi ta ieb_wo_gst_num sgb'jahr'
  noi ta schwerbe sgb'jahr'
  noi ta ieb_staat_num sgb'jahr'
  noi ta hoechsterab sgb'jahr'
  noi ta ieb_sex_id sgb'jahr'
  noi ta alter2007 sgb'jahr'
}
```

```
log close
```


Literaturverzeichnis

- [1] William H. Greene. Econometric Analysis Fifth Edition. International Edition
- [2] STATA. Release 9. References K-Q
- [3] STATA. Release 9. References R-Z
- [4] STATA. Release 9. References Programming
- [5] J.Janssen/W.Laatz. Statistische Datenanalyse mit SPSS für Windows. 6., neu bearbeitete und erweiterte Auflage
- [6] Textausgabe SGB 2009 Mit angrenzenden Gesetzen, Verordnungen und BA-Regeln. Bundesagentur für Arbeit.
- [7] Jörg Heining. FDZDatenreport No. 3/2009
- [8] Merkblatt für Arbeitslose. Bundesagentur für Arbeit
- [9] Paneldaten kategoriale Zielvariablen. http://eswf.uni-koeln.de/lehre/08/3kausalanalyse/ss0803_07.pdf Abruf: 28.07.2010
- [10] Paneldaten kontinuierliche Zielvariablen. http://eswf.uni-koeln.de/lehre/08/3kausalanalyse/ss0803_06.pdf Abruf: 28.07.2010
- [11] Ludwig Fahrmeir Thomas Kneib Stefan Lang. Regression - Modelle, Methoden und Anwendungen. 2. Auflage
- [12] Das Modell der ordinalen Regression. http://www.wiso.uni-hamburg.de/fileadmin/sozialoekonomie/spss-buch/Ergaenzungen/Ordinale_Regression.pdf Abruf: 28.07.2010
- [13] <http://www.uni-bonn.de/~kutikal/anawidaSS06/lec12/Cox-Snell\%20und\%20Nagelkerke.htm>
- [14] Philipp Ronicke. Sozialleistungen.info. <http://www.sozialleistungen.info>, Abruf: 03.05.2010
- [15] Sozialrecht. <http://www.joppo.de/recht/sozialrecht/arbeitslosengeld-i/anwartschaftszeit.html>, Abruf: 01.07.2010

- [16] Bundesrepublik Deutschland. Gesetze im Internet. <http://www.gesetze-im-internet.de>, Abruf: 03.05.2010
- [17] IAB. Benutzerhandbuch IEB. Version 08.00.05
- [18] Sabine Fromm. Binäre logistische Regressionsanalyse. Eine Einführung für Sozialwissenschaftler mit SPSS für Windows. Bamberger Beiträge zur empirischen Sozialforschung Nr. 11, 2005. ISSN 1611-58311. http://www.uni-bamberg.de/fileadmin/uni/fakultaeten/sowi/_lehrstuehle/empirische/_sozialforschung/pdf/bambergerbeitraege/Log-Reg-BBES.pdf, Abruf: 05.05.2010
- [19] <http://www.viavia.ch/ratgeber/pmwiki.php?n=ALV.ZumutbareArbeit>
- [20] Zwei Jahre Datenlieferungen der zugelassenen kommunalen Träger nach § 51b SGB II über XSozial-BA-SGB II http://www.pub.arbeitsagentur.de/hst/services/statistik/000200/html/sgb2/datenlieferungen_zkt_2005_2006.pdf
- [21] http://www.soz.uni-jena.de/adl/m3_regr/html/m3logregr.htm
- [22] Kohler Kreuter. Datenanalyse mit Stata 2.Auflage.
- [23] Glossar der Fernuni Hagen <http://psychologie.fernuni-hagen.de/lernstats/ls/glossar/Dichotomien.html>, Abruf: 10.06.2010
- [24] <http://de.wikipedia.org/wiki/\%C3\%9Cberanpassung>
- [25] Helmut Kopka. \LaTeX . Band 1-3
- [26] http://de.saswiki.org/images/9/9d/13.KSFE-2009-Altenburg-Analyse_von_ROC-Kurven.pdf

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Stellen, die wörtlich oder sinngemäß aus Quellen entnommen wurden, sind als solche kenntlich gemacht.

Diese Arbeit wurde in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegt.

Mittweida, 31.07.2010